

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:03:15 ; Search time 21 Seconds  
(without alignments)  
892.995 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110

Sequence: 1 MESWGLPCLAFPLCFIMAR.....SYFKLNRRNCRTHPEPV 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:1\*  
1: PIR1:1\*  
2: PIR2:1\*  
3: PIR3:1\*  
4: PIR4:1\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1004.5	90.5	180	2	SBX791
2	160.5	14.5	185	2	A60592
3	146	13.2	415	2	F89994
4	139	12.5	527	2	B70700
5	136.5	12.3	435	2	T15143
6	135.5	12.2	300	2	UQ2220
7	128	11.5	348	1	OZQ2BK
8	127.5	11.5	707	2	A46302
9	125	11.3	332	1	OZQ2BK
10	125	11.3	1585	1	MMBEH6
11	124	11.2	3164	1	MMBEH6
12	123.5	11.1	177	1	S65780
13	123.5	11.1	300	2	S19560
14	123.5	11.1	378	2	S19560
15	122	11.0	301	2	E29149
16	121.5	10.9	407	2	T21956
17	121.5	10.9	463	2	T10015
18	121.5	10.9	488	2	F86911
19	121.5	10.9	633	2	F84564
20	121	10.9	676	2	S41022
21	120	10.8	350	2	I55214
22	118.5	10.7	478	2	T21144
23	117	10.5	350	2	T36611
24	117	10.5	922	2	S15204
25	116.5	10.5	437	2	T26767
26	116.5	10.5	500	2	T20961
27	116.5	10.5	530	2	T48627
28	116.5	10.5	940	2	J80291
29	116	10.5	245	1	MAWLB5

## ALIGNMENTS

30	116	10.5	900	2	B70694	probable inf - My
31	115.5	10.4	161	2	S12246	another-specific pr
32	115	10.4	264	2	A54330	major prion protei
33	115	10.4	289	2	F96770	protein RNA-bindin
34	115	10.4	545	2	T15667	hypothetical prote
35	115	10.4	924	2	F87103	initiation factor
36	114.5	10.3	228	2	S46955	microfilament shea
37	114.5	10.3	387	2	T21370	hypothetical prote
38	114.5	10.3	505	2	T50815	cdc2-like protein
39	114.5	10.3	641	2	JC6017	zinc-finger transc
40	114.5	10.3	798	2	T21369	hypothetical prote
41	114.5	10.3	2715	2	T13049	eyelid - fruit fly
42	114	10.3	571	2	T43456	hypothetical prote
43	114	10.3	730	2	T43317	pgl-1 protein - Ca
44	114	10.3	771	2	T29177	hypothetical prote
45	113.5	10.2	260	2	S22373	proline-rich prote

## RESULT 1

SBX791  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence revision 10-Nov-1995 #text\_change 07-May-1999  
C:Accession: S43791  
R:Elis, N.A.; Ye, T.Z.; Patton, S.; German, J.; Goodfellow, P.N.; Weller, P.  
Nature Genet. 6, 394-399, 1994  
A>Title: Cloning of PBDX, an MTC2-related gene that spans the pseudautosomal boundary  
A:Reference number: S43791, MUID:94332149, PMID:8054981  
A:Accession: S43791  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-180 <Full>  
A:Cross-references: GB:573261, NID:G639633, PID:G639634

Query Match 90.5%, Score 1004.5, DB 2, Length 180;  
Best Local Similarity 92.3%; Pred. No. 6.5e-66;  
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY	1	MESWGLPCLAFPLCFIMAR	QY	1	MESWGLPCLAFPLCFIMAR
DB	1	MESWGLPCLAFPLCFIMAR	DB	1	MESWGLPCLAFPLCFIMAR
QY	61	SGGNIYPRKPRPOPOGNSGSGYFNDVDRDGRYP	QY	61	SGGNIYPRKPRPOPOGNSGSGYFNDVDRDGRYP
DB	61	SGGNIYPRKPRPOPOGNSGSGYFNDVDRDGRYP	DB	61	SGGNIYPRKPRPOPOGNSGSGYFNDVDRDGRYP
QY	121	DNTGRGGYRPNRSGYNTYGGDHSITVGNP	QY	121	DNTGRGGYRPNRSGYNTYGGDHSITVGNP
DB	121	DNTGRGGYRPNRSGYNTYGGDHSITVGNP	DB	121	DNTGRGGYRPNRSGYNTYGGDHSITVGNP
QY	181	NNRRNCRTHPEPV 195	QY	181	NNRRNCRTHPEPV 195
DB	166	NNRRNCRTHPEPV 180	DB	166	NNRRNCRTHPEPV 180

## RESULT 2

A60592  
T-cell surface glycoprotein E2 precursor - human  
N:Alternate names: 32k protein, MTC2 antigen  
C:Species: Homo sapiens (man)  
C:Date: 17-Apr-1993 #sequence revision 03-Feb-1994 #text\_change 05-Nov-1999  
C:Accession: S06786, A32876, A60592  
R:Geil, C.; Abbrit, F.; Phallipon, A.; Raynal, B.; Cole, S.; Kaczorek, M.; Bernard, A.  
EMBO J. 8, 3253-3259, 1989  
A>Title: The E2 antigen, a 32 kd glycoprotein involved in T-cell adhesion processes, is  
A:Reference number: S06786, MUID:90059916, PMID:2479542  
A:Accession: S06786  
A:Molecule type: mRNA  
A:Residues: 1-185 <Full>  
A:Cross-references: EMBL:X16996, NID:G30948, PIDN:CAA34863.1, PID:G30949

Query Match	14.5%;	Score 160.5;	DB 2;	Length 185;
Best Local Similarity	28.0%;	Pred. No. 6.2e-05;		
Matches	56;	Conservative	57;	Indels 71;
				Gaps 8

```

OY 10 LAFLCF-----LMHARGQRDPDLADLADDEPFTKKNSDLYPKRKPYY-----53
Db 7 LALLFLGLLGVAAAPDDGGPDLSDLPNE-NKKEPTA-----TPKPSAGDDELDGAVVD 61
OY 54 -----POENPDSGNITYPRPKRRPOPOPGNSGNGYFNDVDRDGRYPRPRPRPAG 108
Db 62 GENDDPRPNP-----PKRMNPNP-NHPSSGSFSDDLADG-----VSG 101
OY 109 GGGGGSSYGSNDNTHGAGYRPNRSYGNTYGGDHSSTYGNDEGNMVAIVSPISVVVV 168
Db 102 GEGKSGSGGGSHRKEGGEADAFG-----VPGIGAVVV 136
OY 169 TLGAAASYFKLNNRRNCFR 188
Db 137 AVAGAISSFIAVQKKCLCPK 156

```

RESULT 3  
 F89994  
 hypothetical protein SA1839 [imported] - *Staphylococcus aureus* (strain N315)  
 C1Species: *Staphylococcus aureus*  
 C1Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C1Accession: F89994  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguni, A.; Mizutani-H, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
 Lancet 357, 1225-1240, 2001  
 A1Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
 A1Reference number: AB9758; MUID:21311952; PMID:11418146  
 A1Accession: F89994  
 A1Status: preliminary  
 A1Molecule type: DNA  
 A1Residues: 1-815 <KUR>  
 A1Cross-References: GB:BA000018; PTD:g13701826; PIDN:BBP43119.1; GSPDB:GN00149  
 A1Experimental source: strain N315  
 C1Genetics:  
 A1Gene: SA1839

```

Query Match      13.2%  Score 146;  DB 2;  Length 415;
Best Local Similarity 27.5%;  Pred. No. 0.0015;
Matches 46;  Conservative 19;  Mismatches 58;  Indels 44;  Gaps 7
QY      33  DPEPRKKPSDLYPKPKP--PYRPOE-----PPDSGNTYRPKRPQPOPGNSGNSGY 86
      :||| ||: | ||||| | |: ||: | |||| | :| |

```

RESULT 4  
B70700  
hypothetical protein Rv0020c - Mycobacterium tuberculosis (strain H37Rv)

hypochlorite protein Rv0020c - Mycobacterium tuberculosis (strain H37RV)  
C1Species: Mycobacterium tuberculosis  
C1Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text\_change 22-Oct-1999  
C1Accession: B70700  
R1Cole, S.T.; Broesch, R.; Parhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon,  
; Connor, R.; Davies, R.; Devlin, K.; Pelletell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajadurai, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A1Authors: Squares, R.; Sultison, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A1Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A1Reference number: A70500; MUID:98295987; PMID:9634230

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-527 <COL>  
A:Cross-references: GB:Z80233; GB:AL123456; NID:g3261645; PID:CA802440.1; PID:c266937  
A:Experimental source: Strain H37Kv  
C:Genetics:  
C:Gene: RV0020c

Query Match	12.5%	Score 139;	DB 2;	Length 527;
Best Local Similarity	29.4%	Pred. No. 0.0062;		
Matches	48;	Conservative 15;	Mismatches 68;	Indels 32; Gaps 9

  

QY	18	HARGRDFDLALDLPETTKKNSDIYK----	PKRYUYQPEPNPDGSGNIYR-----	68
Db	179	VARPD---PRGSRDPGSGSDPRSGYPRFVGGRPRGQGYRRPRNPGRD--	YREQIGYR	233
QY	69	-----PKRP RPORP GNSGSGYFND----	YDRDDGRYPRPRPRPRAG--GGGGGYSYTG	118
Db	234	DQGGYPRGGRYPRGGRYPRDGRGYPDQGGGGRYPRYEDRPRVPSDGRAGYCAPG		233
QY	119	NSDNTHGRGGRYRNSRYGNT--YGEHNSHYG-----	NPEGIMV	155
Db	294	YDQGYRQSGGYRPSRPGGQPGYTG--YGEYGRGRAPRNEGGSYV		334

RESULT 5  
T15143  
hypothetical protein T28P2.8 - *Caenorhabditis elegans*  
C|Species: *Caenorhabditis elegans*  
C|Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
C|Accession: T15143  
R|Madsen, C.; Fromick, B.  
submitted to the EMBL Data Library, April 1997  
A|Description: The sequence of C. elegans cosmid T28P2.  
A|Reference number: Z18300  
A|Accession: T15143  
A|Status: preliminary; translated from GB/EMBL/DBJ  
A|Molecule type: DNA  
A|Residues: 1-635 <MAD>  
A|Cross-references: EMBL:AP000198, NID:g2047345, PID:g2047353, PIDN:AA053055.1, GSFDB:G28P2  
A|Experimental source: strain Bristol N2; clone T28P2

A:Gene: CESP:T28P2.8  
A:Map position: 1  
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8  
Query Match 12.3%; Score 136.5; DB 2; Length 435  
Best Local Similarity 37.9%; Pred. No. 0.0078;

	Matches	50	Conservative	3	Mismatches	54	Indels	25	Gaps	8
QY		33	DPEP-----TKFENSLIYKPK-PPTYPQEPEND-SGNIITPRKRPPOP-----OPNGSG	81						
Db		234	DGPAPAFPTNTVNSPGGGPPAGPPGPFPDGDGGGAQPDPGPppgpgndGPGCGPG	293						
QY		82	NSGGVFND-VRRDDRRYPFRPRPRPPAGGGGGYSYGSNDSMTHGRGGYRPNRSRYGNTRYG	140						
Db		294	QPGEGQGQCGGTDAATCPCP-FRTTPAGGGGGGDFPAGG-----GGGGI-----STGG	340						
QY		141	GDHSHSYGNPEG	152						
Db		341	GGGRADSGAAG	352						
<b>RESULT 6</b>										
JQ2220 hydroxyproline-rich glycoprotein precursor - Chlamydomonas reinhardtii										
C Species: Chlamydomonas reinhardtii										
C Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 21-Jul-2000										
C Accession: U02220										
R Walfenrichmidt, S.; Woessner, J.P.; Beer, K.; Goodenough, U.W.										
Plant Cell 5, 809-820, 1993										
A Title: Isodityrosine cross-linking mediates insolubilization of cell walls in chlamydo-										
mycelium.										
Accession: X02370										
Release number: U02220; MUID:93372571; PMID:7689882										

A.Molecular type: mRNA  
A.Residues: 1-300 <MAP>  
A.Cross-references: GS:L66461; NID:G2898960; PIDD:AA02923.1; PID:G289961  
C.Experimental source: cDNA: VSP-1  
C.Comment: This protein is a component of the insoluble inner layer W2 of vegetative/gamete  
C.Keywords: signal wall, glycoprotein  
E.1-88/Domain: signal sequence #status predicted <Sig>  
E.189-300/Product: hydroxyproline-rich glycoprotein #status predicted <Mat>  
E.189-189,197-205,211-212,222-241/Region: 3-residue repeats (Y-G-G)

```

Query Match      12.2%, Score 135.5, DB 2; Length 300;
Best Local Similarity 30.7%; Pred. No. 0.0064;
Matches 39; Conservative 7; Mismatches 40; Indels 41; Gaps 7;

QY      32 DDEPTTKKNSDIYRKPRPYRPPENPDGSGNTYRPKPPOP-----GNNGN 82
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      142 DSPSPSPPEPVEDSPSPSP--PMDESP-----APESSPVPSPETPPSPAFTYGASPS 193

QY      83 SGGVFNDVDRDDCGYYP-----RPREPPAGGGGGSYSGNSDTHRGGRPNRSRY 135
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      194 PSPFYGGI----GNTGPSVPTYGSSPSPAFPYGGYGGYG-----GYGY----- 237

QY      136 GNTYGGD 142
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      238 -GSYGDD 243

```

```

RESULT 7
OZ2OBK
Circumsporozoite protein precursor - Plasmodium berghei (strain ANKA clone 2.34L)
A:Alternate names: sporozoite surface antigen
C:Species: Plasmodium berghei
C:Date: 30-Jun-1991 #sequence_benchmark
C:Accession: S07873, S12571
R:Lockyer, M.J.; Davies, C.S.; Subhler, A.; Sinden, R.E.
Nucleic Acids Res. 18, 376, 1990
A:Title: Nucleotide sequence of the Plasmodium berghei circumsporozoite protein gene
A:Reference number: S07873; PMID:2183186
A:Accession: S07873
A:Molecule type: DNA
A:Residues: 1-348 <LOC>
A:Cross-references: EMBL:X17606
R:Lockyer, M.J.
submitted to the EMBL Data Library, November 1989
A:Reference number: S12571
A:Accession: S12571
A:Molecule type: DNA

```

A:Residues: -59,'I', 61-81,83-348 <LOC2>  
A:Cross-references: EMBL:X17606; NID:99784; P1D:99785  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: tandem repeat  
F:1-30/Domain: signal sequence #status predicted <STR>  
F:21-348/Product: circumsporozoite protein #status predicted <MAT>  
F:24-205/Region: 8residue repeats  
F:215-247/Region: 2-residue repeats  
F:274-326/Domain: thrombospondin type 1 repeat homology <THR1>

[illegible]

RESULT 8

A46302

PTB-associated splicing factor, long form - human

N:Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding protein

C:Species: Homo sapiens (man)

C:Date: 21-Sep-1993 #sequence, revision 18-Nov-1994 #text\_change 24-Sep-1999

C:Accession: A46302; A43557; S29995

C:Patron, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.

Genes Dev. 7, 393-406, 1993

A:Title: Cloning and characterization of PSP, a novel pre-mRNA splicing factor.

A:Reference number: A46302; PMID:93194059; PMID:8449401

A:Accession: A46302

A:Molecule type: mRNA

A:Residues: 1-707 <P>T>

A:Cross-references: EMBL:X70944; NID:G38457; PIDN:CAA50283.1; PID:G38458

A:Note: sequence extracted from NCBI backbone (NCBIP:127206)

R:Gower, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.

Development 105, 723-731, 1989

A:Title: Cloning and characterization of a myoblast cell surface antigen defined by 24.1.1

A:Reference number: A43557; PMID:90091812; PMID:2480877

A:Accession: A43557

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 312-707 <GOW>

A:Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712

C:Genetics:

A:Gene: GDB:SFPQ; PSF

A:Cross-references: GDB:138275

A:Map position: 4q-4q

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

C:Keywords: alternative splicing; pre-mRNA splicing; surface antigen

C:Keywords: domain: ribonucleoprotein repeat homology <RNM1>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM2>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM3>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM4>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM5>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM6>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM7>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM8>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM9>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM10>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM11>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM12>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM13>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM14>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM15>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM16>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM17>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM18>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM19>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM20>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM21>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM22>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM23>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM24>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM25>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM26>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM27>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM28>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM29>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM30>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM31>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM32>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM33>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM34>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM35>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM36>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM37>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM38>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM39>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM40>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM41>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM42>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM43>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM44>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM45>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM46>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM47>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM48>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM49>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM50>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM51>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM52>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM53>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM54>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM55>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM56>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM57>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM58>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM59>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM60>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM61>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM62>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM63>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM64>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM65>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM66>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM67>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM68>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM69>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM70>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM71>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM72>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM73>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM74>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM75>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM76>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM77>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM78>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM79>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM80>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM81>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM82>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM83>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM84>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM85>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM86>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM87>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM88>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM89>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM90>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM91>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM92>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM93>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM94>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM95>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM96>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM97>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM98>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM99>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM100>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM101>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM102>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM103>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM104>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM105>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM106>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM107>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM108>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM109>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM110>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM111>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM112>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM113>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM114>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM115>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM116>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM117>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM118>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM119>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM120>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM121>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM122>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM123>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM124>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM125>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM126>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM127>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM128>

C:Keywords: domain: ribonucleoprotein repeat homology <RNM12

Query Match	11.5%	Score 127.5	DB 2	Length 707
Best Local Similarity	28.6%	Pred. No. 0.056		
Matches	42	Conservative	5	Mismatches 57; Indels 43; Gaps 6
QY	34	PEPTKKKNSDI-----YPKKPPYRPPENPNPSGSGN-----IYRPKRPQPOPGNSG	81	
DB	147	PGPTPTPPPAVTAAPGAPPPPTPPSSGVPTTPQAQAGPPPPPAVAVPGGEGFGGFPGG	206	
QY	82	NSGGYENDVDGDRYPPRPRPRPPAGGGGGYSGNSDNTHTGRGYRPNRSRYGNTYGG	141	
DB	207	PKG-----GKMPGGPKP-----GGPGLSTGHPKPPRHGGGEGPRG-----GR	245	
QY	142	DHSITY-----GNPEGNMVAKI	158	
DB	246	QHHPPTTHQHHQAPPPPGGRRSEKI	272	



Query Match 11.1%; Score 123.5; DB 2; Length 177;  
 Best Local Similarity 31.9%; Pred. No. 0.028;  
 Matches 45; Conservative 4; Mismatches 43; Indels 49; Gaps 9;

45 YPK---PKPPYPPQENPDGSGNIPKPKPPRPPQPGNSGSGYFNDVDDDRPPRP 101  
 16 YPPAGYPPPGAYPPAGYPPQG---YF-PPPGAYPPAG-----YPPGAYPPAP 58  
 102 RPPRPPAGGGG---GYSSYSGNSDNTGRCGYRNSRYGNTYGGDHSYGNPQNNVA 156  
 59 GGYPPAPGCGYPPAPGCGYPPAP---GHGCGYPPAG---YPA-HHSGAGGICGMIA 109  
 157 KIVSEPIVSVVVTLLGAASY 177  
 110 -----GAAAY 115

## RESULT 13

S19560

proline-rich protein MP4 - mouse

C/Species: Mus musculus (house mouse)

C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Aug-1999

C/Accession: S19560; S22570

R:Roberts, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.

A/Title: Gene sequence of mouse B-type proline-rich protein MP4. Transcriptional start F

A/Reference number: S19560; MUID:92111548; PMID:1765104

A/Accession: S19560

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-300 &lt;ROB&gt;

A/Cross-references: GB:X58438; NID:953181; PTDN:CAA1344.1; PID:953182

R:Roberts, S.G.E.; Layfield, R.; McDonald, C.J.

Nucleic Acids Res. 19, 5205-5211, 1991

A/Title: The mouse proline-rich protein MP6 promoter binds isoprenaline-inducible parotid

A/Reference number: S22570; MUID:92020206; PMID:1747160

A/Accession: S22570

A/Molecule type: DNA

A/Residues: 1-14 &lt;RO2&gt;

A/Cross-references: EMBL:X61126

C/Superfamily: proline-rich protein

Query Match 11.1%; Score 123.5; DB 2; Length 300;  
 Best Local Similarity 31.2%; Pred. No. 0.047;  
 Matches 53; Conservative 6; Mismatches 58; Indels 53; Gaps 11;

18 HARGQRFDLADLDEPTKPKNSDIYKPKPPY-----POPEN---PDGSGNIV 66  
 93 HAPRPPQ---GDALGPPRP---GNQGGPSPGPPQSSQGRPPQGNQGGPPRPPQGG--- 142  
 67 PRPKRPPQPPGN-----SGNSGYFNDVDRDDGRYPPRPP---R 104  
 143 --PQGRP-PQPGNQGPPPPPGPQGRPPQPGNQG-----PQGGPPRP-PRPGNQGPP 193  
 105 PRAGGGGGGYSYSGNSDNTGRCGYRNSRYGNTYGGDHSYGNP-GN 153  
 194 PPQGGPQGRPPQGNQGGPPQGGPQAPPRPGNQGPPPPRPPRGN 243

## RESULT 14

S14959

proline-rich protein - wheat

C/Species: Triticum aestivum (common wheat)

C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 24-Nov-1999

C/Accession: S14959

R:Raines, C.A.; Lloyd, J.C.; Chao, S.; John, U.P.; Murphy, G.J.P.

A/Title: A novel proline-rich protein from wheat.

A/Reference number: S14959; MUID:91329699; PMID:1714320

A/Accession: S14959

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-378 <RAI>  
 A/Cross-references: EMBL:X52472; NID:921841; PTDN:CAA36712.1; PID:921842  
 C/Superfamily: hydroxyproline-rich glycoprotein

Query Match 11.1%; Score 123.5; DB 2; Length 378;  
 Best Local Similarity 28.6%; Pred. No. 0.059;  
 Matches 24; Conservative 15; Mismatches 28; Indels 17; Gaps 2;  
 29 DALDDEPTKPKNSDIYKPKPPYPPQPE---NPDGSGNIPKPKPPRPPQPGNSGSG 85  
 302 EPMKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPE 353  
 86 YFNDVDRDDGRYPPRPPRPPAG 109  
 354 -----KPEPKPEPKPEPKPEPKG 371

## RESULT 15

E29149

proline-rich protein - mouse

C/Species: Mus musculus (house mouse)

C/Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 20-Aug-1999

C/Accession: E29149

R:Clements, S.; Mehanho, H.; Carlson, D.M.

J. Biol. Chem. 260, 13471-13477, 1985

A/Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence

A/Reference number: A92501; MUID:86033799; PMID:3840480

A/Accession: E29149

A/Status: Clone pUM125

A/Molecule type: mRNA

A/Residues: 1-301 &lt;CLE&gt;

A/Cross-references: GB:M1897; NID:9200540; PTDN:AAA40001.1; PID:9200541

C/Superfamily: proline-rich protein

Query Match 11.0%; Score 122; DB 2; Length 301;  
 Best Local Similarity 28.8%; Pred. No. 0.06;  
 Matches 45; Conservative 9; Mismatches 52; Indels 50; Gaps 7;

34 PEPTKPKNSDIYKPKPPYPPQENPDGSGNIPKPKPPQ---PQGNNS----- 80  
 148 PQGNQGG-----PPPGGPPQGRPPQGNQGGPPPGGPPRPPGPPPGGPPQ 201  
 81 -----GNSGYFNDVDRDDGRYPPR-----PPR-----PPAGGGGGYS 115  
 202 RPTQPGNQG---PQGGGPPGPPRPGNQGCPPPGGPPGPPRPGNQGPPGPGGPP 258  
 116 SYGNSDNTGRCGYRNSRYGNTYGGDHSYGNP 151  
 259 RPPGNGGPPPGGPPRPPQGNHGGPPRH---GNNE 291

Search completed: February 18, 2004, 19:07:04  
 Job time: 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 18, 2004, 18:56:46 ; Search time 17 Seconds

(without alignments)  
539,424 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110  
Sequence: 1 MESHWGLPCALPCPLMHAR.....SYFKLNRRNCFTHEPENY 195

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1004.5	90.5	180	1 XG_HUMAN	P55808 homo sapien
2	160.5	14.5	185	1 MIC2_HUMAN	P14209 homo sapien
3	128	11.5	347	1 CSP_PLABA	P23093 plasmodium
4	127.5	11.5	707	1 SPPQ_HUMAN	P23246 homo sapien
5	125	11.3	339	1 CSP_PLABE	P06915 plasmodium
6	124.5	11.2	1183	1 DRPL_RAT	P54258 rattus norv
7	124	11.2	3164	1 TEGU_HSV1	P10220 herpes simp
8	119	10.7	245	1 VEA_HPV05	P06924 human papil
9	118	10.6	1790	1 SPRA_EMENI	P78621 emericella
10	117	10.5	922	1 PERT_BORPA	P24328 bordetella
11	116	10.5	245	1 VEA_HPV5B	P26550 human papil
12	116	10.5	900	1 IP2_MYCTU	P16163 mycobacteri
13	115.5	10.4	161	1 ASP1_HELAN	P22357 helianthus
14	115	10.4	264	1 PRIO_BOVIN	P10279 bos taurus
15	115	10.4	924	1 IF2_MYCIE	P02519 mycobacteri
16	113.5	10.2	1664	1 SIPI_GLOTM	P06852 clostridium
17	113	10.2	824	1 TGM1_RAT	P23606 rattus norv
18	112.5	10.1	255	1 PRIO_CANPA	P06501 canis fami
19	112.5	10.1	704	1 RP3A_BOVIN	P06846 bos taurus
20	112	10.1	304	1 VEA_HPV47	P22421 human papil
21	112	10.1	1059	1 CAPU_DROME	P24120 drosophila
22	111.5	10.0	448	1 AAC2_DICTI	P74196 dictyostella
23	110.5	10.0	143	1 PAR1_TRYBB	P74196 dictyostella
24	110.5	10.0	145	1 PARC_TRYBB	P06084 trypanosoma
25	110.5	10.0	257	1 PRIO_MUSPP	P52114 mus muscu
26	110.5	10.0	257	1 PRIO_MUSPP	P52114 mus muscu
27	110.5	10.0	283	1 EXTN_SORBI	P40244 mus muscu
28	110.5	10.0	345	1 SOD_DROME	P08473 drosophila
29	110	9.9	911	1 IIF3_MOUSE	P21814 mus muscu
30	109.5	9.9	911	1 COTT_BACSU	P11863 bacillus su
31	109.5	9.9	449	1 APG_BRANA	P40603 brassica na
32	109	9.8	256	1 PRP2_BOVIN	P01880 bos taurus
33	109	9.8	305	1 ROAD_HUMAN	Q13151 homo sapien

34	109	9.8	373	1 RO31_XENLA	P51668 xenopus lae
35	108.5	9.8	129	1 PARB_TRYBB	P09791 trypanosoma
36	108.5	9.8	174	1 PRPB_HUMAN	P81469 homo sapien
37	108.5	9.8	256	1 PRIO_CERRE	P79142 ceruus elap
38	108.5	9.8	526	1 FUS_HUMAN	P35637 homo sapien
39	108.5	9.8	534	1 AFG_ARATH	P40602 arabidopsis
40	108.5	9.8	625	1 DUS8_HUMAN	Q13202 homo sapien
41	108	9.7	256	1 PRIO_CAPI	P52113 capra hircu
42	108	9.7	256	1 PRIO_FELCA	O18754 felis silve
43	108	9.7	256	1 PRIO_SHEEP	P23907 ovis aries
44	108	9.7	514	1 MEPP_MOUSE	P63943 mus muscu
45	107.5	9.7	512	1 FUS_BOVIN	Q28009 bos taurus

## ALIGNMENTS

RESULT 1  
XG\_HUMAN  
ID XG\_HUMAN STANDARD; PRT; 180 AA.  
AC P55808;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE XG glycoprotein precursor (Protein PBDX).  
GN XG OR PBDX.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=94332149; Pubmed=8054981;  
RA Ellis N.A., Ye T.Z., Patton S., German J., Goodfellow P.N.,  
RA "Cloning of PBDX, an MIC2-related gene that spans the pseudoautosomal  
RT boundary on chromosome Xp.";  
RL Nat. Genet. 6:394-400(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Moko S., Fujiki K., Kanai A., Tanaka Y., Iwata T.,  
RT "Identification of PBDX gene highly expressed in human cornea.";  
RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -1- PTM: O-Glycosylated (PROBABLE).  
CC -1- POLYMORPHISM: XG IS RESPONSIBLE FOR THE XG BLOOD GROUP SYSTEM.  
CC -1- SIMILARITY: TO PROTEIN MIC2/CD99.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC  
CC EMBL; X96421; -; NOT ANNOTATED CDS.  
DR EMBL; AP380356; AAL04055.1; -;  
DR PIR; S43791; S43791.  
DR GeneW; HGNC:12806; XG.  
DR MIM; 314700; -;  
KW Transmembrane; Glycoprotein; Blood group antigen; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 180 XG GLYCOPROTEIN.  
FT DOMAIN 22 142 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 143 163 POTENTIAL.  
FT DOMAIN 164 180 CYTOPLASMIC (POTENTIAL).  
SO SOURCE 180 AA; 19723 MW; DADA95B859C4530 CRC64;  
Query Match 90.5%; Score 1004.5; DB 1; Length 180;  
Best Local Similarity 92.3%; Pred. No. 4,3e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

Qy	1	MESMWGPGICLAFICPLMHARGORPDLAALADDEPTKKKPNSDYPRKPRPYQPENPD	60
Db	1	MESMWGPGICLAFICPLMHARGORPDLAALADDEPTKKKPNSDYPRKPRPYQPENPD	60
Qy	61	SGGNIVRPPRRQPPQPGNSGNSGGYFNDVDRDGRYPFRPRPRPPAGGGGYSYGS	120
Db	61	SGGNIVRPPRRQPPQPGNSGNSGGYFNDVDRDGRYPFRPRPRPPAGGGGYSYGS	120
Qy	121	DNTHGRGCGYRPNRSRYGNTTGGDHSITYGNEGNNVAKIVSPVSVVVTLLGAASYFKL	180
Db	121	DNTH-----GGDHSITYGNEGNNVAKIVSPVSVVVTLLGAASYFKL	180
Qy	181	NNRRNCRFTHPEPNV	195
Db	166	NNRRNCRFTHPEPNV	180

RESULT 2

ID	MIC2	HUMAN	STANDARD	PTI	185	AA
AC	P14209	000518				
DT	01-JAN-1990	(Rel. 13, Created)				
DT	01-JAN-1990	(Rel. 13, Last sequence update)				
DT	28-FEB-2003	(Rel. 41, Last annotation update)				
DE	T-cell surface glycoprotein E2 precursor (E2 antigen) (CD99 antigen) (MIC2 protein) (1257)					
DE	(CD99X OR CD99 OR MIC2X OR MIC2) AND (CD99Y OR CD99 OR MIC2Y OR MIC2).					
GN	Hom sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 23-39.					
RC	TISSUE=T-cell;					
RA	MEDLINE=90059916; PubMed=2479542;					
RA	Gelin C., Aubrit F., Phallion A., Raynal B., Cole S., Kaczorek M., Bernard A.;					
RT	"The E2 antigen, a 32 kd glycoprotein involved in T-cell adhesion processes, is the MIC2 gene product."					
RT	EMBO J. 8:3253-3259(1989).					
RL	[2]					
RN	SEQUENCE FROM N.A. (ISOFORM II).					
RP	Park S.H., Hahn J.H., Kim M.K., Sohn H.W., Choi E.Y., Kim S.H.;					
RT	"An alternative splicing form of CD99 (MIC2).";					
RL	Submitted (DEC-1996) to the EMBL/GenBank/DBS databases.					
RN	[3]					
RP	SEQUENCE OF 10-105 FROM N.A.					
RP	MEDLINE=87217010; PubMed=3472717;					
RA	Darling S.M., Goodfellow P.J., Pym B., Banting G.S., Pritchard C., Goodfellow P.N.;					
RA	"Molecular genetics of MIC2: a gene shared by the human X and Y chromosomes";					
RL	Cold Spring Harb. Symp. Quant. Biol. 51:205-212(1986).					
RN	[4]					
RP	SEQUENCE OF 1-22 FROM N.A.					
RP	MEDLINE=88289765; PubMed=2456574;					
RA	Goodfellow P.J., Mondello C., Darling S.M., Pym B., Little P., Goodfellow P.N.;					
RT	"Absence of methylation of a CpG-rich region at the 5' end of the MIC2 gene on the active X, the inactive X, and the Y chromosome.";					
RL	Proc. Natl. Acad. Sci. U.S.A. 85:5605-5609(1988).					
CC	-1- FUNCTION: INVOLVED IN T-CELL ADHESION PROCESSES. IT IS INVOLVED IN SPONTANEOUS ROSETTE FORMATION WITH ERYTHROCYTES.					
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).					
CC	-1- ALTERNATIVE PRODUCTS:					
CC	Event=Alternative splicing; Named isoforms=2;					
CC	Name=I;					
CC	Isoid=P14209-1; Sequence=Displayed;					
CC	Name=II;					
CC	Isoid=P14209-2; Sequence=VSP_004324;					
CC	-1- PTM: EXTENSIVELY O-GLYCOSYLATED.					
CC	-1- SIMILARITY: TO PROTEIN XG/PBDX.					

```
-- This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcement/or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X16986; CA034863.1; -
DR EMBL; U082164; AAB58501.1; -
DR DR EMBL; M16279; AAA02999.1; -
DR DR EMBL; J03841; AAAS9848.1; -
DR PIR; S06786; A60592.
DR DR Genbank; HGNC:7082; CD99.
DR MIM; 313470; -
DR MIM; 450000; -
DR GO; GO:0005737; Cytoplasm; TAS.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005888; C:protein integral to plasma membrane; TAS.
KW T-cell; Glycoprotein; Cell adhesion; Transmembrane; Signal; Alternative splicing; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 185
FT TRANSMEM 123 147
FT VARSPPLIC 159 185
FT FT
FT VARIANT 166 166 /PRTID=VAR_014733.
FT FT VARIANT 173 173 N->I (IN DBSNP:4717) .
FT SQ SEQUENCE 185 AA; 18948 MW; C30E09EBB022EAB CRC64;
Query Match 14.5%; Score 160.5; DB 1; Length 185;
Best Local Similarity 28.0%; Pred. No. 0.0001;
Matches 56; Conservative 16; Mismatches 57; Indels 71; Gaps 8;
QY 10 LAFLCF---LMHARGRDFLDALDDPEPTKKPSNDLYPKPPYY----- 53
DB 7 LALLPLGLTLGVLVAAPDGGFSLDALPDNE-NKRPFA----IPKPAGADFLLGDVVVD 61
QY 54 -----PQEPNDSGNITPRPKPPQPQGSGNSGVFNVDNRDPKRRPRPPAG 108
DB 62 GENDDPRPNP-----PKPPNPNN-PHPSSGSFSADLADG-----VSG 101
QY 109 GGCGGYSYSGNSDNTHTRGGRGYPMSRYGNTYGCDHSHTYGNPEGNMVAKIVSPIVSVVV 168
DB 102 GEKGSGSDGGSHKKEBEADApg-----VIPIGVAVVV 136
QY 169 TLCAAAASYFKLNRRNRCFR 188
DB : :::::
DB 137 AVAGAISFFIAYOKKCLCFK 156
QY : :::::
DB : :::::
RESULT 3
CSP PLABA STANDARD; PRT; 347 AA.
AC P23093;
DT 01-NOV-1991 (Rel. 20, Created)
DT DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Circumsporozoite protein precursor (CS).
OS OS Plasmodium berghei (strain Anka).
OC OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX OX NCBI_TaxID=5823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90221834; PubMed=2183186;
RA Lockyer M.J., Davies C.S., Subrtler A., Sinden R.B.;
RT Nucleotide sequence of the Plasmodium berghei circumsporozoite protein gene from the ANKA clone 2.34L." ;
RL Nucleic Acids Res. 18:376-376(1990).
```

```

CC -1- FUNCTION: THE CIRCUMSPOROZITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOROZITE (THE INJECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -1- SIMILARITY: Contains 1 TSP type-1 domain.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; X17606; CAA35608.1; -.
CC DR PIR; S07873; OZQ0BK.
CC DR InterPro; IPR003067; Circmsprzoite.
CC DR InterPro; IPR000884; TSP1.
CC DR Pfam; PF00090; Tsp_1; 1.
CC DR PRINTS; PR01303; CIRCMSPRZOITE.
CC DR SMART; SM00209; TSP1, 1.
CC DR PROSITE; PS50092; TSP1, 1.
CC KW Malaria; Sporozoite; Repeat; Signal.
CC FT SIGNAL 1 23 POTENTIAL.
CC FT CHAIN 24 347 CIRCUMSPOROZITE PROTEIN.
CC FT DOMAIN 93 204 13 X 8 AA REPEATS.
CC FT DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.
CC FT DOMAIN 274 325 TSP TYPE-1.
CC SQ SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;
CC -----
QY Query Match 11.5%; Score 128; DB 1; Length 347;
Db Best Local Similarity 30.8%; Pred. No. 0.03;
QY Matches 32; Conservative 15; Mismatches 33; Indels 24; Gaps 4
QY 32 DDEPTKKKNSDIPPKRP-----PYYPQENP-DGGNTYPPKRPQPPQGN 80
Db 171 NDPPP---PNANDPPPPNDPPAPNPANDPPPPPNNDPAPQGNNNPQPPRPQPP--- 224
QY 81 GNSGCGYFNDVDGRYPYRPRPRPPAGGGGGGYSYSGNSDNTH 124
Db 225 -----QPPQPPQPPQPPQPPRPPQPPQPPGGGNNNNKNNNNNDSTY 261
-----
RESULT 4
SFPO_HUMAN
ID ID STANDARD; PRT; 707 AA.
AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Splicing factor, proline and glutamine-rich (Polypyrimidine tract-
DE binding protein-associated splicing factor) (PMB-associated splicing
DE factor) (PSF) (DNA-binding p52/Plp100 complex, 100 kDa subunit).
DE SFPO OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=93194059; PubMed=8449401;
RA Paton J.G., Porro E.B., Galceean J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor.";
RL Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Fetal skeletal muscle;
RX MEDLINE=90091812; PubMed=2480877;
RX

```

```

RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayk R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RL defined by 24.1D5 monoclonal antibody.";
RN Development 105:723-731(1989).
[3]
RP SEQUENCE OF 48-68 AND 213-246.
RX MEDLINE=93176127; PubMed=8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT "Purification and characterization of a DNA-binding heterodimer of 52
and 100 kDa from HeLa cells.";
RL Biochem. J. 290:267-272(1993).
CC -I- FUNCTION: ESSENTIAL PRE-mRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPliceosome FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
CC -I- SUBUNIT: HETEROTRIMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Long;
CC IsoId=P23246-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P23246-2; Sequence=VSP_005655;
CC -I- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -I- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
CC ECTOKINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X70944; CAA50283.1; -.
CC DR EMBL; X16850; CAA34747.1; -.
CC DR PIR; A46302; A46302.
CC DR HSSP; P11940; 1CVJ.
CC DR SWISS-2DPAGE; P23246; HUMAN.
CC GeneW; HGNC:10774; SFPQ.
CC DR GK; P23246; -.
CC DR MI; G05199; -.
CC DR GO; G00008248; F:Pre-mRNA splicing factor activity; TMS.
CC DR GO; G00006371; P:mRNA splicing; TMS.
CC DR InterPro; IPR000504; RNA_rec_mot.
CC Pfam; PF00076; rrm; 2.
CC SMART; SMO0360; RRM; 2.
CC DR PROSITE; PS50102; RRM; 2.
CC DR PROSITE; PS00030; RRM_RNP_1; 1.
CC KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
CC Alternative splicing.
FT DOMAIN 297 369 RRM-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 9 11 1.
FT REPEAT 19 21 2.
FT REPEAT 25 27 GLN/GLU/PRO-RICH.
FT DOMAIN 10 266 POLY-GLY.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-ARG.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VAAPSLIC 663 707 RTTFEGGGGAGPYGGGPGTPAGYGRGREYRGPNK
KKPF->YMIIVG (in isoform Short).

```



```

PT  CONFLICT 243 243 /Ptid=VSP_005855.
SQ  SEQUENCE 707 AA; 76149 MW; 6080D5A95E255847 CRC64; G -> R (IN REF. 3).

Query Match 11.5%; Score 127.5; DB 1; Length 707;
Best Local Similarity 28.6%; Pred. No. 0.065;
Matches 42; Conservative 5; Mismatches 57; Indels 43; Gaps 6;

Qy 34 PEPTKKPMSDI-----YKPKRPYPPQENPDGSGN-----IYPRKPRPQPPQNSG 81
Dy 147 PGPPTPPPAVTSAPPGAPPPPTBSGGVPTTPQAAGPPPPAAVPGGPBKGGPPGG 206
Dy 82 NSGGFYNDVDDDDGRRPPRRPPRPPAGCGGGGYSYSGNSDNTGGGGRPPSRGNTYGG 141
Dy 207 PKG-----GMPGCGPR-----CGCGGLRPGCHPRPRRGSGPRG-----GR 245
Qy 142 DHSTY-----GNPBGNGYAKI 158
Dy 246 QHHPPYHQHHQGGPPGPGGRSBEKI 272

RESULT 5
CSP_PLABE STANDARD; PRT; 339 AA.
ID_CSP_PLABE
AC 066915;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Circumsporozoite protein precursor (CS).
OS Plasmodium berghei.
NCBI_TaxID=5821;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=87089740; PubMed=2432395;
RT Eichinger D.J., Arnot D.E., Tam J.P., Nussenzeiwig V., Enea V.;
RL "Circumsporozoite protein of Plasmodium berghei: gene cloning and
RL identification of the immunodominant epitopes.";
RL Mol. Cell. Biol. 6:3965-3972(1986).
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VETERINATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE PARASITISM.
CC -1- SIMILARITY: Contains 1 TSP type-1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M14135; AAA29577.1; -.
DR PIR; A44948; OZQMB.
DR InterPro; IPR003067; Circmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF000090; tap_1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SM00209; TSP1_1.
DR PROSITE; PS50092; TSP1_1.
DR Malaria; Sporozoite; Repeat; Signal.
FT CHAIN 1 23 PROBABLE.
FT DOMAIN 24 339 CIRCUMSPOROZOITE PROTEIN.
FT DOMAIN 93 196 13 X 8 AA TANDEM REPEATS.
FT DOMAIN 206 238 16 X 2 AA TANDEM REPEATS OP P-Q.
FT DOMAIN 286 317 TSP TYPE-1.
SQ SEQUENCE 339 AA; 37138 MW; E8068A6D11D9551B CRC64;

Query Match 11.3%; Score 125; DB 1; Length 339;

```

```

Beet Local Similarity   32.3%, Pred. No.0.048;
Matches      32, Conservative    15; Mismatches     30; Indels       22; Gaps        5;

Oy      32 DDEPTKKNPNDIYPRKKPYRY--POEENPDS-----GQNIYPRKPFPQPQGSGNSGCG 85
Db      171 NDPAF---PVAN-----DPAFPNANDPPPNPNVDPAAPPGQNNNPQPQPPRPQP----- 216

Oy      86 YFNVDNRDDGRYPFRPRRPPAGGGGGSSYSNGSDMTW 124
Db      217 --GFQPPQPPQPPQPPRPPQPPQPPQPCGGNNNNKNNNNNDSTY 253

RESULT 6
DRPL RAT
ID DRPL RAT STANDARD; PRT; 1193 AA.
AC PS4258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DN Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=917396;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
RT homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Bpielen J.T., Riess O.;
RA "Predominant neuronal expression of the gene responsible for
RL dentatorubral-pallidoluysian atrophy (DRPLA) in rat.";
Hum. Mol. Genet. 4:1619-1624(1995).
CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/sib.ch).
CC or send an email to licenses@sib-sib.ch).
CC -----
CC EMBL; U31777; AAA8037.1; -.
CC EMBL; X89453; CA61623.1; -.
CC InterPro; IPR002951; Atrophin.
CC Pfam; PF03154; Atrophin-1; 2.
CC PRINTS; PR01222; ATPPHIN.
CC FT DOMAIN 165 171 POLY-PRO.
CC FT 303 306 POLY-PRO.
CC FT 377 383 POLY-SER.
CC FT 387 391 POLY-SER.
CC FT 440 446 POLY-SER.
CC PT DOMAIN 477 480 POLY-PRO.
CC PT 480 POLY-HIS.
CC FT DOMAIN 481 489 POLY-GLN.
CC FT 502 505 POLY-PRO.
CC FT DOMAIN 562 572 POLY-SER.
CC FT 702 705 POLY-PRO.
CC FT 705 POLY-PRO.
CC FT CONFLICT 455 455 N -> S (IN REF. 2).
CC FT CONFLICT 594 594 F -> L (IN REF. 2).
CC FT CONFLICT 594 594 F -> R (IN REF. 2).
CC FT CONFLICT 689 689

```



CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eurotiiales; Trichocomaceae; Emmentella.  
 OK NCBI\_TaxID=162425;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97361829; PubMed=9218790;  
 RA Harris S.D., Hamer L., Sharpless K.E., Hamer J.E.;  
 RT "The *Aspergillus nidulans* *sepa* gene encodes an Fhl/2 protein involved  
 RT in cytokinesis and the maintenance of cellular polarity.";  
 RL EMO 1.6:3474-3483(1997).  
 RN [2]  
 RP REVISIONS TO 143-153: 207; 1071-1109 AND 1644.  
 RA Hamer L., Harris S.D., Sharpless K.E., Hamer J.E.;  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 559-1790 FROM N.A., AND FUNCTION.  
 RC STRAIN=FGSC 26;  
 RX MEDLINE=95229045; PubMed=7713416;  
 RA Mathoul J.F., Adams T.H.;  
 RT "Identification of developmental regulatory genes in *Aspergillus*  
 RT *nidulans* by overexpression.";  
 RL Genetics 139:537-547(1995).  
 CC -1- FUNCTION: INVOLVED IN CYTOKINESIS. OVEREXPRESSION RESULTS IN  
 CC GROWTH INHIBITION.  
 CC -1- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE  
 CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.  
 CC -1- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.  
 CC -1- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.  
 CC -1- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.  
 CC -1- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.  
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. EN11  
 CC SUPFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U83658; AA663335.3; -; ALT\_SEQ.  
 DR EMBL, U83658; AA663335.3; -; ALT\_SEQ.  
 DR EMBL, U83658; AA663335.3; -; ALT\_SEQ.  
 DR InterPro: IPR003104; FH2; 1.  
 DR Pfam: PF02181; FH2; 1.  
 DR SMART: SM00498; FH2; 1.  
 DR Cell division: Coiled coil.  
 CC KW  
 FT DOMAIN 258 486 GBD.  
 FT DOMAIN 376 718 FH3.  
 FT DOMAIN 724 811 COILED COIL (POTENTIAL).  
 FT DOMAIN 955 1136 FH1 (PRO-RICH).  
 FT DOMAIN 1141 1658 FH2.  
 FT DOMAIN 1435 1566 COILED COIL (POTENTIAL).  
 FT DOMAIN 1586 1600 DAD.  
 FT DOMAIN 1608 1611 ANG/DYS-RICH (BASIC).  
 FT CONFLICT 1071 1109 PPPPPGFGGPPPPPPGFGGPPPPPPPGGARG  
 FT -> LSPSSSTGFWMTTPAASFTWFWMTSTAASAAAYWVIC  
 FT C (IN REF. 3).  
 FT D -> V (IN REF. 3).  
 FT CONFLICT 1476 1476 V -> L (IN REF. 3).  
 FT CONFLICT 1504 1504 V -> L (IN REF. 3).  
 SQ SEQUENCE 1790 AA; 197355 MW; 192136DE2EF2A75B CRC64;  
 CC -----  
 CC Query Match 10.6%; Score 118; DB 1; Length 1790;  
 CC Best local similarity 35.2%; Pred. No. 0.69;  
 CC Matches 32; Conservative 0; Mismatches 37; Indels 22; Gaps 3;  
 CC -----  
 CC 26 DLADALDDEPTKKN-----SDIYKPKPYTPQENPDGSGNTYPRKRPQPGNSG 81  
 CC 1010 DDATAAPPPEPPPPPPHPSGAAAPPP-----PPPPPPGAGAAPPPPPPPPPGGIG 1066  
 CC 82 NSGGYFNDVDRDGRYPRPRPPPPAGGGG 112

Db 1067 G-----PPPPPPPPGFGG 1082  
 RESULT 10  
 ID PERT BORPA STANDARD; PRT; 922 AA.  
 AC P24328;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Pertactin precursor (outer membrane protein P.70) (P.95).  
 GN PRN.  
 OS Bordetella parapertussis.  
 CC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 CC Alcaligenaceae; Bordetella.  
 OK NCBI\_TaxID=519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CN2591;  
 RX MEDLINE=91251771; PubMed=2041476;  
 RA Li L.J., Dougan G., Novotny P., Charles I.G.;  
 RT "P.70 pertactin, an outer-membrane protein from *Bordetella*  
 RT *parapertussis*: cloning, nucleotide sequence and surface expression in  
 RT *Bacteriella coli*.";  
 RL Mol. Microbiol. 5:409-417(1991).  
 CC -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS  
 CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN  
 CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.  
 CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)  
 CC CONCENTRATIONS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, X54547; CA38419.1; -;  
 DR EMBL, X54547; CA38419.1; -;  
 DR EMBL, A26124; CA01786.1; -;  
 DR EMBL, A19182; CA01454.1; -;  
 DR PIR, S15204; S15204.  
 DR InterPro: IPR006315; Autotransporter.  
 DR InterPro: IPR005546; Autotransporter.  
 DR InterPro: IPR004899; Pertactin.  
 DR InterPro: IPR003991; Pertactin\_C.  
 DR InterPro: IPR003992; Pertactin\_N.  
 DR Pfam: PF03797; Autotransporter; 1.  
 DR Pfam: PF03212; Pertactin; 1.  
 DR PRINTS: PRO1482; PERTACTIN.  
 DR PRINTS: PRO1484; PERTACTIN.FAMILY.  
 DR TIGRFAMs: TIGR01414; autotransp\_bar1; 1.  
 KW Outer membrane; Signal; Virulence; Repeat.  
 FT SIGNAL 1 34  
 FT CHAIN 35 922 P.95.  
 FT PROPEP 648 922 POTENTIAL.  
 FT SITE 260 262 PERTACTIN (P.70).  
 FT DOMAIN 266 290 CELL ATTACHMENT SITE (INVOLVED IN  
 FT REPEAT 266 270 ADHESION TO VARIOUS EUKARYOTIC CELL  
 FT REPEAT 271 275 LINES).  
 FT REPEAT 276 280 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.  
 FT REPEAT 281 285 2.  
 FT DOMAIN 575 603 3.  
 FT SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;  
 CC -----  
 CC 4 (APPROXIMATE).  
 CC 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.  
 CC 3DF7BC58D4712478 CRC64;

IP2_MYCTU	STANDARD;	PRT;	900 AA.
ID	IP2_MYCTU		
AC	P71613;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Translation initiation factor IPF-2.		
GN	INFB OR RV2839C OR MT2905 OR MTCY1657.03.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Coriobacteriales; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Badcock S.V., Eiglmister K., Gas S., Barry C.E. III, Tekala R.,		
RA	Davies R., Baahm D., Brown D., Chillingworth T., Connor R.,		
RA	Holmesby T., Jagals K., Kirogh A., Mclean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajendream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC 1551 / Oshkosh;		
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Petersen J., Deboy R., Dodson R., Gilm M.L., Hatt D., Hickey E.,		
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,		
RA	Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikita A.,		
RA	Bishei W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains."		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: One of the essential components for the initiation of		
CC	protein synthesis. Protects formylmethionyl-tRNA from spontaneous		
CC	hydrolysis and promotes its binding to the 30S ribosomal subunit.		
CC	Also involved in the hydrolysis of GTP during the formation of the		
CC	70S ribosomal complex. (By similarity).		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-1- SIMILARITY: BELONGS TO THE IPF-2 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.lab-sib.ch/announce/">http://www.lab-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL; Z81331; CAB03670.1; -.		
DR	EMBL; AE007115; AAK47231.1; -.		
DR	PIR; B70694; B70694.		
DR	HSSP; P13551; IDAR.		
DR	TIGR; MT2905; -.		
DR	Tuberculist; RV2839C; -.		
DR	HAMAP; MF_00100; - 1.		
DR	InterPro; IPRO00795; EF_Grpbind.		
DR	InterPro; IPRO04161; EFTU_D2.		
DR	InterPro; IPRO00178; IF2_.		
DR	InterPro; IPRO06847; IF2_N.		
DR	InterPro; IPRO01806; Ras_tnsfmg.		
DR	InterPro; IPRO05225; Small_GRP.		
DR	Pfam; PF00009; GTP_EFTU; 1.		
DR	Pfam; PF03144; GTP_EFTU_D2; 2.		
DR	Pfam; PF04760; IF2_N; 2_.		
DR	PRINTS; PR00315; ELONGATINFCT.		
DR	PRINTS; PR00449; RASTRNSFRMG.		
DR	ProDom; PD186100; IF2; 1.		

DR ProdOm; PD002594; G\_Purothionin; 1.

RI bovine spongiform encephalopathy.  
 RL J. Infect. Dis. 167:602-613(1993).  
 RN [4]  
 RP SEQUENCE FROM N. A.

RI bovine spongiform encephalopathy.  
 RL J. Infect. Dis. 167:602-613(1993).  
 RN [4]  
 RP SEQUENCE FROM N. A.

CC STRAIN=Holstein-Friesian; TISSUE=Brain;  
RA Horiuchi M.,  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN (5)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Jersey;  
RX MEDLINE=21422903; PubMed=11531705;  
RA Hill D., Comincini S., Schaefer J., Dolf G., Ferretti L.,  
RA Williams J.L.,  
RT "Complete genomic sequence of the bovine prion gene (PRNP) and  
RT polymorphism in its promoter region.";  
RL Anim. Genet. 32:231-232(2001).  
RN (6)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Korean;  
RX Yoo H.S., Kang S.G., Choi I.S., Kang S.K., Hwang W.S.,  
RA "Nucleotide sequence of Prp cDNA in Korean cattle.";  
RT Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
RN (7)  
RP SEQUENCE OF 1-15 FROM N.A.  
RA Tanaka M., Inoue S., Ikeda T., Horiuchi M., Ishiguro N., Shinagawa M.,  
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
RN (8)  
RP SEQUENCE OF 25-36.  
RX MEDLINE=89057122; PubMed=2904126;  
RA Hope J., Reekie L.J.D., Hunter N., Multhaup G., Beyreuther K.,  
RA White H., Scott A.C., Stack M.J., Dawson M., Wells G.A.,  
RT "Fibrils from brains of cows with new cattle disease contain scrapie-  
RT associated protein.";  
RL Nature 336:390-392(1988).  
RN (9)  
RP STRUCTURE BY NMR OF 132-241.  
RX MEDLINE=20359707; PubMed=10899999;  
RA Lopez Garcia F., Zahn R., Riek R., Muehlich K.,  
RT "NMR structure of the bovine prion protein.";  
RT Proc. Natl. Acad. Sci. U.S.A. 97:8334-8339(2000).  
RN (10)  
RP FUNCTION: THE FUNCTION OF PRP IS NOT KNOWN. PRP IS ENCODED IN THE  
RA HOST GENOME AND IS EXPRESSED BOTH IN NORMAL AND INFECTED CELLS.  
CC -1- SUBUNIT: PRP HAS A TENDENCY TO AGGREGATE YIELDING POLYMERS CALLED  
CC "RODS".  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS  
CC INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,  
CC CRETZFELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME  
CC (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),  
CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.  
CC -1- SIMILARITY: BELONGS TO THE PRION FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; X55882; CAA39368.1; -  
DR EMBL; D10612; BAA01467.1; -  
DR EMBL; D10613; BAA01468.1; -  
DR EMBL; S55629; AAB25514.1; -  
DR EMBL; AB001468; BAA19253.1; -  
DR EMBL; AF517842; AAW66709.1; -  
DR EMBL; D26151; BAA05138.1; -  
DR PIR; A54330; A54330.  
DR PDB; 1DWY; 26-FEB-02.  
DR PDB; 1DWZ; 26-FEB-02.  
DR PDB; 1DXO; 26-FEB-02.  
DR PDB; 1DX1; 26-FEB-02.  
DR InterPro; IPR000817; Prion.  
DR Pfam; PF03377; Prion; 1.  
DR Pfam; PF03991; Prion octapep; 6.  
DR PRINTS; PR00341; PRION.

DR SMART; SM00157; PRP; 1.  
DR PROSITE; PS00291; PRION 1; 1.  
DR PROSITE; PS00706; PRION 2; 1.  
KW Prion; Brain; Glycoprotein; GPI-anchor; Repeat; Signal; Polymorphism;  
KW 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 25 264  
FT CARBOHYD 192 192  
FT CARBOHYD 208 208  
FT DISULFID 190 225  
FT DOMAIN 54 103  
FT REPEAT 54 62  
FT REPEAT 63 70  
FT REPEAT 71 78  
FT REPEAT 79 86  
FT REPEAT 87 94  
FT REPEAT 95 103  
FT VARIANT 71 78  
FT CONFLICT 218 218  
FT STRAND 140 141  
FT STRAND 140 141  
FT HELIX 155 162  
FT HELIX 165 167  
FT STRAND 173 174  
FT HELIX 184 203  
FT TURN 204 206  
FT TURN 211 237  
SQ SEQUENCE 264 AA; 28614 MW; D6D21403816A231 CRC64;  
Query Match 10.4%; Score 115; DB 1; Length 264;  
Best Local Similarity 32.9%; Pred. No. 0.18;  
Matches 28; Conservative 9; Mismatches 32; Indels 16; Gaps 3;  
QY 72 RQPGSGNSGSGYFNDVDRDGRYPFRFR-----RPAGGGGGSYSGSDMTHRG 127  
DB 25 KRPFGGGGWNNG-----SRYPGSGGNGRYPFGGGGNGGPGGGGPGGG 75  
QY 128 GYFNSRYGNTYGGDHSYGNFEG 152  
DB 76 WQGF---HGGWGQPHGGWGQPHG 97  
RESULT 15  
ID IF2 MYCLE STANDARD; FRT; 924 AA.  
AC 092519;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DE Translation initiation factor IF-2.  
GN INFB OR ML1556 OR MLCB596.14.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eigmeister K., Fairhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthey S., Fellwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornby T., Jagele K., Lacroix C., Maclean J., Moulé S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellern J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 409:1007-1011(2001).  
CC -1- FUNCTION: One of the essential components for the initiation of

```

CC protein synthesis. Processes formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunit.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE IP-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL035472; CAB836570.1; -.
CC EMBL: AL5853922; CAC30507.1; -.
CC PIR: P87103; P87103.
CC HSSP: P13551; 1ELO.
CC Leptoma; ML1556; -.
CC DR HAMAP: MP_00100; -. 1.
CC DR InterPro: IPR000795; BP_GTPbind.
CC DR InterPro: IPR004161; EFTU_D2.
CC DR InterPro: IPR000178; IP2.
CC DR InterPro: IPR006847; IP2_N.
CC DR InterPro: IPR005225; Sma11_GTP.
CC DR Pfam: PP00009; GTP_EFTU_1.
CC DR Pfam: PP03144; GTP_EFTU_D2; 2.
CC DR Pfam: PP04760; IP2_N_2.
CC DR PRINTS: PR00315; ELONGATNFCT.
CC DR ProDom: PD186100; IP2; 1.
CC DR TIGRFAMs: TIGR00487; IP-2; 1.
CC DR TIGRFAMs: TIGR00231; sma11_GTP; 1.
CC DR PROSITE: PS01176; IP2; PALSE_NMG.
CC KW Initiation factor; Protein biosynthesis; GTP-binding;
CC FT Complete proteome.
CC KW DOMAIN 423 575 G-DOMAIN.
CC FT NP_BIND 429 436 GTP (BY SIMILARITY).
CC FT NP_BIND 479 483 GTP (BY SIMILARITY).
CC FT NP_BIND 533 536 GTP (BY SIMILARITY).
CC FT DOMAIN 216 219 POLY-GLY.
CC FT DOMAIN 271 279 POLY-GLY.
CC SQ SEQUENCE 924 AA; 96650 MW; 34605A1069001224 CRC64;
CC -----
Cc Query Match 10.4%; Score 115; DB 1; Length 924;
Cc Best Local Similarity 31.0%; Pred. No. 0.59;
Cc Matches 44; Conservative 3; Mismatches 49; Indels 46; Gaps 8;
Cc -----
Cc 46 PKPKPPYPOQ--PKNPDGSGNITPPKPRP-----QPPGNSGSGYENVND 91
Cc ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cc Db PAPKGGKPPRVGNPFFSSAOSVAPRIPRPAPRPSASPSMSPRGAVGGG-----P 220
Cc -----
Cc 92 RDDGGRPPRP---RRPPAGG---GGGGYS-----YGNSDNTHGR-----GGYR 130
Cc ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cc Db RPPRTGVRRPGGGRGAPVGGASDAGGNGYRGVGAALPGGSGGGRGPGGGGGGGR 280
Cc -----
Cc 131 PMSRYGNTYGGDHHSTYGNPBG 152
Cc ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cc QY PGQR-----GGAAGAPRRGG 296
Cc -----
Cc Db

```

Search completed: February 18, 2004, 19:05:43  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:02:25 ; Search time 35 Seconds

(without alignment)  
1437.722 Million cell updates/sec

Title: US-09-997-701-1

Sequence: 1 MESMGGLPLAFLCLMHR.....STFKLNRRNCRTPEPV 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriap:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	16.3	175	11	Q8VCN6
2	167.5	15.1	174	11	Q8BLN1
3	165.5	14.9	157	6	Q46670
4	164.5	14.8	173	6	Q02788
5	149.5	13.5	213	4	Q8TC21
6	149.5	13.5	419	16	Q8NVK9
7	149	13.4	262	4	Q8TC22
8	147	13.2	134	11	Q8C1L8
9	146	13.2	235	13	Q8QFW9
10	146	13.2	415	16	Q8QFW9
11	143	12.9	354	10	Q8FYE4
12	142	12.8	1066	10	Q8H093
13	141	12.7	190	4	Q8TC20
14	141	12.7	521	16	Q8VKT1
15	140	12.6	461	13	Q8AX89
16	139	12.5	527	16	P71590

17	138.5	12.5	246	11	Q8R1R5	Q8R1R5 ratius norv
18	138.5	12.5	374	11	Q9JUL7	Q9JUL7 ratius norv
19	135.5	12.2	300	10	Q07373	Q07373 chlamydomon
20	135	12.2	381	10	Q9FKA5	Q9FKA5 arabidopsis
21	135	12.2	381	10	Q8H7G9	Q8H7G9 arabidopsis
22	135	12.2	594	5	Q9VEP4	Q9VEP4 drosophila
23	134.5	12.1	108	5	Q8GYX8	Q8GYX8 faecicola he
24	134.5	12.1	880	16	Q8F7K1	Q8F7K1 leptospira
25	133	12.0	735	3	Q9UVQ3	Q9UVQ3 podospira a
26	132.5	11.9	623	11	Q8K2U8	Q8K2U8 mus musculu
27	132	11.9	237	11	Q8BIF0	Q8BIF0 mus musculu
28	132	11.9	460	5	Q9VFZ6	Q9VFZ6 drosophila
29	132	11.9	595	10	Q8S1W9	Q8S1W9 oryza sativ
30	130.5	11.8	164	10	Q8LDL1	Q8LDL1 saccharum h
31	130.5	11.8	2344	5	Q9N3Y8	Q9N3Y8 caenorhabdi
32	130	11.7	463	5	Q93535	Q93535 caenorhabdi
33	127.5	11.5	247	10	Q8H5X2	Q8H5X2 oryza sativ
34	127.5	11.5	247	10	Q8H5W8	Q8H5W8 oryza sativ
35	127.5	11.5	637	4	Q8BSV4	Q8BSV4 homo sapien
36	127.5	11.5	872	11	Q55000	Q55000 ratius norv
37	127	11.4	497	5	Q27512	Q27512 caenorhabdi
38	125.5	11.3	358	10	Q8W265	Q8W265 oryza sativ
39	125.5	11.3	374	2	Q33899	Q33899 streptococc
40	125.5	11.3	513	10	Q9LV82	Q9LV82 arabidopsis
41	125.5	11.3	513	10	Q8W4P1	Q8W4P1 arabidopsis
42	125.5	11.3	525	11	Q9PB56	Q9PB56 mus musculu
43	125.5	11.3	525	11	Q9PB55	Q9PB55 mus musculu
44	125.5	11.3	525	11	Q9ESB3	Q9ESB3 mus musculu
45	125.5	11.3	1379	5	Q8SWX1	Q8SWX1 drosophila

#### ALIGNMENTS

RESULT 1  
Q8VCN6 PRELIMINARY; PRT; 175 AA.  
ID Q8VCN6  
AC Q8VCN6; 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Similar to RIKEN CDNA 2410026K10 gene.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RA Strausberg R.;  
RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC019482; AAL19482.1; -  
SQ SEQUENCE 175 AA; 16782 MW; A309B6E2494FBE1 CRC64;

Qy	10	LAFLCFMHARGQRFDLADALDDPEPTKKNPSDIYPKKPPYYPQENPDSG-----	62
Db	16	LALIGAAARGAASDDPFGDLEDP-----NMKTPPKA---PTPKKPSGGFLEBDL	64
Qy	63	-----GNTPPKPPKPPQPPQNSGSGYFNDVDRDGRYPKPPKPPKPPAGGG	110
Db	65	PGGGGGGAGKPPGN---PPQDPKPK-PPPHDSGG-IDSIDLADL-----AQGG	108
Qy	111	GGGYSSYNSD--NTHGRGYRPPNSRYGNTYGGDHSTYGNDEGMVAIVSPISVVVV	168
Db	109	GGGAGRRSGDBGGHGGAGABPR-----GTPQG-----LVPGVAAYVA	148
Qy	169	TLGGAASYFKLNNRNCFR	188
Db	149	AVAGAVSSFVAVYQRRRLCFR	168



## RESULT 2

06BLN1 PRELIMINARY; PRT; 174 AA.  
 ID 06BLN1  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical glycolytic region containing protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK044051; BAC1754.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 174 AA; 16719 MW; 580017FD553A4B18 CRC64;

Query Match 15.1%; Score 167.5; DB 11; Length 174;  
 Best Local Similarity 31.5%; Pred. No. 3.5e-07;  
 Matches 57; Conservative 20; Mismatches 73; Indels 31; Gaps 10;

QY 10 LALFLCFLMHARGQDFDLADLDDEPTKPSNDIYKPKPPYY----- 69  
 DB 16 LALLGLGALVLAADDDGFDLSALPEKE-DKKPTA-----TPKKSAGDPPDLGAVVD 60  
 QY 70 KPRPQOPGSGNSGGYFNDVDDGRYPKPPRPP-AGCGGGGYSYSGNSDNTHERGG 128  
 DB 61 ---BDALPGGGGGAGG-----EKPGNR--PQDPKPPPHDSGDISDGLAD--AAQGG 109  
 QY 129 YRPNRYGNTYGGDHHSTYTG-NPFGNNVAKIVSPISVSVVTLTGAAASYFKLNRRNCF 187  
 DB 110 --GAGRGSGDEGGHGGAGGAEPEGTPOG-LVPGVVAALVPPVAGAVSSFAVQRRLCF 166  
 QY 188 R 188  
 DB 167 R 167

## RESULT 3

046670 PRELIMINARY; PRT; 157 AA.  
 ID 046670  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CD99 type II-COS7 (Fragment).  
 OS Cercopithecus aethiops (Green monkey) (Grivet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Cercopithecus.  
 NCBI\_TaxID=9534;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Park S.H.;  
 RT "A monkey homolog of an alternative spliced form of human CD99."  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U82166; AAB93833.1; --  
 FT NON TER 1  
 SQ SEQUENCE 157 AA; 15679 MW; C34DB239B4D55E20 CRC64;

Query Match 14.9%; Score 165.5; DB 6; Length 157;  
 Best Local Similarity 27.1%; Pred. No. 4.6e-07;  
 Matches 55; Conservative 19; Mismatches 56; Indels 71; Gaps 8;

QY 10 LALFLCFLMHARGQDFDLADLDDEPTKPSNDIYKPKPPYY----- 53  
 DB 1 LALLGLGALVLAADDDGFDLSALPEKE-DKKPTA-----TPKKSAGDPPDLGAVVD 55  
 QY 54 -----PQENPDGSGNIYPRPKPPQOPGNSGNSGGYFNDVDDGRYPKPPRPPAG 108  
 DB 56 GGNDPPPPNP-----PKKPNPNPQAQSSGS-FSDADLADG-----VSG 95  
 QY 109 GGGGGYSYSGNSDNTHERGGYRPNRSRYGNTYGGDHHSTYGNPFGNNVAKIVSPISVSVV 168  
 DB 96 GEGKGGSDGGSGPRKGEADAPG-----VTPGIVGAIVV 130  
 QY 169 TLGAAASYFKLNRRNCFRTH 191  
 DB 131 AVAGAISSFLAYOKKLCFRKND 153

## RESULT 4

002788 PRELIMINARY; PRT; 173 AA.  
 ID 002788  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CD99-Cos7 (Fragment).  
 OS Cercopithecus aethiops (Green monkey) (Grivet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Cercopithecus.  
 NCBI\_TaxID=9534;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Park S.H., Hahn J.-H., Kim M.K., Sohn H.W., Choi B.Y., Kim S.H.;  
 RT "A monkey homolog of human CD99."  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U82165; AAB58504.1; --  
 FT NON TER 1  
 SQ SEQUENCE 173 AA; 17299 MW; B9667ABED16287B8 CRC64;

Query Match 14.8%; Score 164.5; DB 6; Length 173;  
 Best Local Similarity 27.2%; Pred. No. 6.3e-07;  
 Matches 55; Conservative 18; Mismatches 58; Indels 71; Gaps 8;

QY 10 LALFLCFLMHARGQDFDLADLDDEPTKPSNDIYKPKPPYY----- 53  
 DB 1 LALLGLGALVLAADDDGFDLSALPEKE-DKKPTA-----TPKKSAGDPPDLGAVVD 55  
 QY 54 -----PQENPDGSGNIYPRPKPPQOPGNSGNSGGYFNDVDDGRYPKPPRPPAG 108  
 DB 56 GGNDPPPPNP-----PKKPNPNPQAQSSGS-FSDADLADG-----VSG 95  
 QY 109 GGGGGYSYSGNSDNTHERGGYRPNRSRYGNTYGGDHHSTYGNPFGNNVAKIVSPISVSVV 168  
 DB 96 GEGKGGSDGGSGPRKGEADAPG-----VTPGIVGAIVV 130  
 QY 169 TLGAAASYFKLNRRNCFRTH 190  
 DB 131 AVAGAISSFLAYOKKLCFRKAN 152

## RESULT 5

08TC21 PRELIMINARY; PRT; 213 AA.  
 ID 08TC21  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE MIC21L isoform E3-84.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Ox	NCB1_TaxID=96606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISUB=Sketetal muscle;
RA	Suh Y.H., Shin Y.K., Park S.H.;
RT	"MIC2L1, a novel human paralog of CD99 (MIC2), B3-E4 isoform.";
RL	Submitted (FEB-2002) to the EMBL/genbank/DDbJ databases.
DR	EMBL: AY078166; AAL86618.1; ..
SO	SEQUENCE 213 AA; 22833 MW; P90595SDAA9BA7D5 CRC64;
Query Match	13.5%; Score 149.5; DB 4; Length 213;
Best Local Similarity	27.1%; Pred. No. 1.7e-05;
Matches	55; Conservative 29; Mismatches 80; Indels 39; Gaps 9
Oy	1 MESWGG-LPCLAFCLFMHARGQDFD--LADALDDPEPTKPKNSDIYPRKPYPYQ 55
Dd	1 MVAMRSATLVCIASFSLATLVGRSGDPDFDLDAVKETSVKGQMNHYTTTKAPVTR 60
Oy	56 PENPDGSGNIYPRKPRPOPOPQNGSNGSGGYND-VDRDDGAYPRPRPRPPAGGGGGY 114
Dd	61 -----APANTLGNDPFLDALDDLDRNDDDR-----RKPIAGGGG-- 95
Oy	115 SSYGNSDNTHRGGRGPRPSRYGN-TYGGDHSTHYGNPGSNVAK--IYSIVSVVVITLL 171
Dd	96 FSDKDLEIYVGGEYKPKDKGDKDGYYGN----DPSGMVAEFGTTAGVASALMALLI 150
Oy	172 GAASVFPLNNRNCERTHEPEN 194
Dd	151 GAVSSYSISYQQKKCFCSIQGGIN 173
RESULT 6	
OBNVK9	PRELIMINARY; PRT; 419 AA.
AC	OBNVK9;
DT	01-OCT-2002 (TREMBLrel. 22. Created)
DT	01-OCT-2002 (TREMBLrel. 22. Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23. Last annotation update)
DE	MW1956 protein.
GN	Staphylococcus aureus (strain MW2).
OS	Bacteria; Firmicutes; Bacillales; Staphylococcus.
OC	NCBI_TaxID=196620;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=22040717; PubMed=12044378;
RA	Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA	Nagai Y., Iwano N., Asano K., Naito T., Kuroda H., Cui L.,
RA	Yamamoto K., Hiratake S.K.;
RT	"Genome and virulence determinants of high virulence community-
RT	acquired MRSA.";
RL	Lancet 359:1819-1827 (2002).
RL	EMBL: AP004829; BAB95821.1; -
DR	InterPro: IPRO02965; P rich extensin.
DR	PRINTS; PRO1217; PRICHEXTENSIN.
KW	Complete proteome.
SO	SEQUENCE 419 AA; 46630 MW; E31BD7F6525A1B18 CRC64;
Query Match	13.5%; Score 149.5; DB 16; Length 419;
Best Local Similarity	27.0%; Pred. No. 3.5e-05;
Matches	44; Conservative 17; Mismatches 55; Indels 47; Gaps 6;
Oy	34 PEPTKPNDSIYPRKPYPYQPEKPDGSGGIYF---RPKRPOPOPNGSNGSGGYFDV 90
Dd	126 PKPDPKDPD-KPRNPDPKPDPNKKNPKPKDPDKPKKNPKPKP-----DP 173
Oy	91 DRDDGRIPRRRRPP-----AGGGGGYSGSYGNSDNTHRGGRYENSR 134
Dd	174 DKPRNPDPKDPKPKPNPNPSPDPDQPDGSNHSGSKNGTWNPAASDGNMGQMQRGN 233
Oy	135 YGNYYGGDHSTYGNPGSNVAKIYSPITVSIVVVITLLGAASY 177
Dd	234 QGNSO-----NPTGN-----DFVSGRFALANGARKY 260

ID	Q8TC22	PRELIMINARY;	PRT;	262 AA.
AC	Q8TC22;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	MIc2L1 isoform E3'-E4'-E3-E4 (MIc2, like 1).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RA	Sub Y.H., Shin Y.K., Park S.H.;			
RT	"MIc2L1, a novel human paralog of CM99 (MIc2), E3'-E4'-E3-E4			
RL	isoform."			
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RA	Strausberg R.;			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY078165; AAL86617.1; -			
DR	EMBL; BC030536; AAK00536.1; -			
DR	Genew; HGNC:18237; MIC2L1.			
SO	SEQUENCE 262 AA; 27986 MW; 3A254961D32C191 CRC64;			
Query Match 13.4%; Score 149; DB 4; Length 262;				
Best Local Similarity 25.4%; Pred. No. 2.3e-05;				
Matches 60; Conservative 29; Mismatches 91; Indels 56; Gaps 11;				
Qy	1 MESWVG--LPCIAFLCFMARGQRDPD--LADALDPPEPTKCP--NSDIYPKKPPY 53			
Db	1 MVAARSAFVICTAFSLATLVQSGDPDFTNLEAVKETSVAQPMHTTTTNNRPGTT 60			
Qy	54 POPENPDSG-----NITYPPKPPPOPOPGNSGSGYFN 88			
Db	61 RAPAKPPGSGGIDLADALDDDDRRKPKGIGGERHNVTTTTTKRPVTTAPANTLGNDFD 120			
Qy	89 --DV-----DRDGRVPRPRPPRPAAGGGGGYSYGNDSNTHGGRGRRNSRYGN--TYG 140			
Db	121 LADLDDRRNDDBDR-----RKPIAGGGG--PSDKDLIDVGGGKPKDCKGGRGYG 171			
Qy	141 GDHSHSTGNEEGNMVAK--IVSPVSVVVVTLTGAASYFEKLNRRRCFRTHEPEN 194			
Db	172 SN-----DDGSGMVAERGTIAGVASLMAALIGAVSYISYQKKCFSTIGGLN 222			
RESULT 8				
ID	Q8C1L8	PRELIMINARY;	PRT;	134 AA.
AC	Q8C1L8;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Hypothetical glycine-rich region containing protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RA	STRAIN=C57BL/6J;			
RA	MEDLINE=22354683; PubMed=12466851;			
RA	the PANTOM Consortium,			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RL	Nature 420:563-573 (2002).			
RL	EMBL; AK010597; BAC25305.1; -			



DR EMBL; AL391716; CAC05499.1; -  
 DR EMBL; AY081342; AL091231.1; -  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; ehand; 2.  
 DR ProDom; PD000012; EF-hand; 1.  
 DR SMART; SM00054; EFh; 2.  
 DR PROSITE; PS00018; EF HAND; 2.  
 SQ SEQUENCE 354 AA; 37113 MW; F827C8342E258EA2 CRC64;

Query Match 12.9%; Score 143; DB 10; Length 354;  
 Best Local Similarity 31.4%; Pred. No. 0.00011;  
 Matches 49; Conservative 15; Mismatches 52; Indels 40; Gaps 8;

QY 36 PTKKPNVDYPRKPPRYPP-----ENPSSGNITPRKPRPQPPQNGSNGGYF 87  
 DB 63 PYYGSSSYAPPPSAPYAPSPGDYKPKPKPYGGG--YGAPP-----SGSS----- 108  
 QY 88 NIVDRDDGHRPPRRPRPPAGGGG-----GYSYSGNDYHGRGYR-NSRYGNTY 139  
 DB 109 -----DYSGYGAQPRPSQSGHGGYGATPRHGVSDYG---SYGAPPSPASSGHGGY 159  
 QY 140 GG-DHSHYTGPNBGNMVAIVSVVVVTLGAA 174  
 DB 160 GGYPQASYSYSPASLIPSGFAPGTDPNIVACFQA 195

RESULT 12  
 ID 08H093 PRELIMINARY; PRT; 1046 AA.

AC 08H093; 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Putative phospholipase.  
 GN OSJNBA009608.23.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhacridae; Oryzae; Oryza.  
 OK NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Ganberger K., Kim M.M.,  
 RA Overton II L.L., Bera J.J., Tsiliris T., Krol M.L., Varrault B.B.,  
 RA Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Vanaken S.S.,  
 RA Utecherback T.T., Feldblum T.V., Yang Q.Q., Haas B.J., Suh B.B.,  
 RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNBA009608 genomic sequence."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC078894; AA00720.1; -  
 SQ SEQUENCE 1046 AA; 116544 MW; BBA3A8C1985EA577 CRC64;

Query Match 12.8%; Score 142; DB 10; Length 1046;  
 Best Local Similarity 25.2%; Pred. No. 0.00045;  
 Matches 51; Conservative 18; Mismatches 67; Indels 66; Gaps 9;

QY 23 RQFDLADALDDEPTKPP-----NSDIYPRKPPRY-----POPENPD-----SG 62  
 DB 48 RSFHCVPASAPRPPQPPQFVAHHSAPLQPPRPPHNAVPPRPPHPSPPVYDPAHAAA 107  
 QY 63 GNIYPRKPRP-----QPPGNSGNSGYPNDV-----DRDGRPPPPR 102  
 DB 108 YPSYSPNSPISPSSSSFHHHPSPSPSPSYSLADGLANMHSRHH--YPPPS 165  
 QY 103 PR-----PPAGGGGGYSGNSDNTHGGRYRPNRYGNTYGGDHS-----T 146  
 DB 166 PAAVPAASPSVLPSPASFPGGSSHG-----CGGMQVYPRPPAGSGHGMQVVA 217  
 QY 147 YGNPBNMVAIVSVIVSVVV 168  
 DB 218 YGSPAGSGHSGVSRSLKVVLL 239

RESULT 13  
 ID 08TC20 PRELIMINARY; PRT; 190 AA.

AC 08TC20; 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Mic2L1 isoform E4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Suh Y.H., Shin Y.K., Park S.H.;  
 RT "Mic2L1, a novel human paralog of CD99 (MIC2), E4 isoform."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY078167; AAL86619.1; -  
 SQ SEQUENCE 190 AA; 20300 MW; F21A43C7F6B319F6 CRC64;

Query Match 12.7%; Score 141; DB 4; Length 190;  
 Best Local Similarity 23.7%; Pred. No. 8.2e-05;  
 Matches 52; Conservative 22; Mismatches 51; Indels 94; Gaps 8;

QY 1 MESWNG--LPCLAPFLCFLMARG-----QRFDLADALDDEPTK 38  
 DB 1 MVAMSAFLVCLAFSLATLVORSGSDPDDFNLEDAVKETSVKRDPLADALDNRN--- 57  
 QY 39 KPNSDIYPRKPPRYPPPENPDSGNTYPRKPRPQPPQNGSNGGYFNDVDRDGRYP 98  
 DB 58 -----DRDGR-- 63  
 QY 99 PRPPRPPAGGGGGYSGNSDNTHGGRYRPNRYGN-TYGGDHSYGNPBNMVAIVAK 157  
 DB 64 -----RKRIAGGGG--PSDKDLEDIYVGGGEYKPKDKGDKGYGSA-----DDPSGVAE 111  
 QY 158 --IVSPISVVVVTLGAAASYFKLNRRNCFRTHEPN 194  
 DB 112 PGTIAGVASALAMALIGAVSYISYQKKCFCSIQGLIN 150

## RESULT 14

08VKT1 PRELIMINARY; PRT; 521 AA.

AC 08VKT1; 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Hypothetical protein MT0023.  
 GN MT0023.

OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OK NCBI\_TaxID=1773;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ernolova M.D., Salzberg S.L.,  
 RA Delcher A., Utecherback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF006916; AAK44245.1; -

DR TIGR; MT0023; -  
 DR InterPro; IPR000253; FHA.  
 DR Pfam; PF00498; FHA; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR PROSITE; PS5006; FHA\_DOMAIN; 1.  
 KW Hypothetical protein.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 18:56:15 ; Search time 40 Seconds

(without alignments)  
773.792 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110  
1 MBSWGLPCIAFLCFLMAR.....SYFKANRRCFTHEPENY 195

Sequence: 1 MBSWGLPCIAFLCFLMAR.....SYFKANRRCFTHEPENY 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1110	100.0	195	21	AAW90961	Human CSGP-1 prote
2	1110	100.0	195	23	ABG73109	Human cell surface
3	885	79.8	168	22	AAW23909	Human EST encoded
4	502	45.2	96	22	ABG00690	Novel human diagno
5	491	44.2	91	21	AAW00621	Human secreted pro
6	465	41.9	345	22	ABG27670	Novel human diagno
7	427	38.5	91	22	ABG27669	Novel human diagno
8	371.5	33.5	73	22	ABG00689	Novel human diagno
9	308	27.7	62	22	ABG03178	Novel human diagno

10	176.5	15.9	169	21	AA19128	polypeptide isolat
11	172.5	15.5	165	23	ABG72314	Rat protein isolat
12	164.5	14.8	195	21	AAW58379	Lung cancer associ
13	164.5	14.8	195	24	ABU04013	Human expressed pr
14	162.5	14.6	265	22	ABG11668	Novel human diagno
15	160.5	14.5	185	24	ABU04008	Human expressed pr
16	160.5	14.5	185	24	ABU04009	Human expressed pr
17	160.5	14.5	185	24	ABU04010	Human expressed pr
18	160.5	14.5	185	24	ABU04011	Human expressed pr
19	160.5	14.5	185	24	ABU04012	Human expressed pr
20	160.5	14.5	185	24	ABU04014	Human expressed pr
21	160.5	14.5	185	24	ABU04015	Human expressed pr
22	149.5	13.5	419	24	ABU19109	Pathogen specific
23	149.5	13.5	423	24	ABU18931	Pathogen specific
24	149	13.4	262	22	AAW5634	Homo sapiens secre
25	149	13.4	262	22	AAW24121	Human EST encoded
26	149	13.4	262	22	AAW38873	Human polypeptide
27	149	13.4	262	22	AAW27243	Human EXMD-21 SEQ
28	149	13.4	274	22	AAW40659	Human polypeptide
29	147.5	13.3	512	22	ABG00692	Novel human diagno
30	145	13.1	528	22	AAW81105	Mycobacterium tube
31	143	12.9	51	22	ABG27668	Novel human diagno
32	143	12.9	354	21	AAW44289	Arabidopsis thalia
33	141.5	12.7	368	21	AAW26449	Arabidopsis thalia
34	139	12.5	527	23	ABU05697	M. tuberculosis an
35	138	12.4	229	20	AAW95625	Homo sapiens secre
36	135	12.2	369	21	AAW47270	Arabidopsis thalia
37	135	12.2	381	21	AAW47292	Arabidopsis thalia
38	135	12.2	594	21	ABG61362	Drosophila melano
39	132.5	11.9	355	22	AAW40927	Human polypeptide
40	132	11.9	460	22	ABG63425	Drosophila melano
41	131	11.8	381	21	AAW26698	Arabidopsis thalia
42	130.5	11.8	163	23	AAW14306	Sugarcane proline
43	127.5	11.5	707	22	AAW39141	Human polypeptide
44	126	11.4	1260	22	ABW59060	Drosophila melano
45	125	11.3	204	20	AAW13498	Tissue cement prot

## ALIGNMENTS

RESULT 1	AAW90961	standard; Protein; 195 AA.
ID	AAW90961;	
XX	AAW90961;	
AC	14-JUL-2000 (first entry)	
XX		
DT		
XX		
DE	Human CSGP-1 protein.	
XX		
KW	Human; CSGP-1; cell surface glycoprotein; CSGP-2; anti-anemic;	
KW	neutropenic; neutrophilic; anticonvulsant; gene therapy; treatment;	
KW	diagnosis; hematologic disorder; anemia; myeloproliferative disorder;	
KW	lymphoma; karyotypic disorder; Klinefelter syndrome; Turner syndrome;	
KW	neural disorder; akathisia; Alzheimer's disease; amnesia; catatonia;	
KW	epilepsy; neurofibromatosis.	
XX		
OS	Homo sapiens.	
XX		
PN	US6043056-A.	
XX		
PD	28-MAR-2000.	
XX		
PF	06-NOV-1998; 98US-0187331.	
XX		
PR	06-NOV-1998; 98US-0187331.	
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Yue H, Gorgone GA, Baughn MR, Corley NC, Guegler KJ;	
XX		
DR	WPI; 2000-270341/23.	

DR N-PSDB; AAIL1721.  
XX Polymucleotide encoding cell surface glycoproteins is useful for  
PT diagnosis, treatment and prevention of hematologic disorders,  
PT karyotypic disorders and neuronal disorders -  
XX  
XX Claim 1, Fig 1A-C; 38pp; English.  
XX  
CC This invention describes the novel human cell surface glycoproteins  
CC CSGP-1 and CSGP-2. The products of the invention have anti-neuritic,  
CC neurotropic, neuroprotective and anticonvulsant activity and can be used  
CC for gene therapy. The encoding nucleic acids are useful for the  
CC diagnosis, treatment and prevention of hematologic disorders such as  
CC anemia, myeloproliferative disorders and lymphoma, karyotypic disorders  
CC such as Klinefelter syndrome and Turner syndrome, and neuronal  
CC disorders such as akathisia, Alzheimer's disease, amnesia, cataplexy,  
CC epilepsy and neurofibromatosis. The products of the invention are also  
CC useful as probes for mapping the gene sequences encoding CSGP-1 and  
CC CSGP-2. This sequence represents the human CSGP-1 protein described in  
CC the method of the invention.  
XX  
SQ Sequence 195 AA;  
Query Match 100.0%; Score 1110; DB 21; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.6e-88;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MESWNGLPCLAFICPLMHARGRDPDLADALDDEPTKKNSDIYPRKPPYPOBNPD 60  
DB 1 MESWNGLPCLAFICPLMHARGRDPDLADALDDEPTKKNSDIYPRKPPYPOBNPD 60  
QY 61 SGNITPRPKRPPQPOGNSGSGYFNDVDRDGRYPYPRRPPRPAAGGGGGYSSYGN 120  
DB 61 SGNITPRPKRPPQPOGNSGSGYFNDVDRDGRYPYPRRPPRPAAGGGGGYSSYGN 120  
QY 121 DNTHGGRGYRPNRSGNTYGGDHSTYGNPEGNVAKIVSIYVVVTLGAAASYFKL 180  
DB 121 DNTHGGRGYRPNRSGNTYGGDHSTYGNPEGNVAKIVSIYVVVTLGAAASYFKL 180  
QY 181 NNRNRCRTHEPENV 195  
DB 181 NNRNRCRTHEPENV 195  
RESULT 2  
ABG73109  
ID ABG73109 standard; Protein, 195 AA.  
XX  
XX ABG73109;  
XX  
XX 04-MAR-2003 (first entry)  
DE Human cell surface glycoprotein (CSGP)-1.  
XX  
XX Human, cell surface glycoprotein; CSGP, CSGP-1; recognition;  
XX adhesion; cell motility; signal transduction; stage-specific marker;  
XX cell differentiation; immunological identification; diagnostic;  
XX gene therapy; vaccine; haematologic disorder; anaemia; haemorrhage;  
XX thrombosis; haematopoietic disorder; haemoglobin disorder;  
XX sickle cell anaemia; bone marrow disorder; leukaemia;  
XX myeloproliferative disorder; lymphoma; Hodgkin's disease; haemophilia;  
XX karyotypic disorder; Klinefelter syndrome; Turner syndrome;  
XX neuronal disorder; Alzheimer's disease; amnesia; dementia; depression;  
XX amyotrophic lateral sclerosis; Down's syndrome; epilepsy; schizophrenia;  
XX Huntington's disease; multiple sclerosis; Parkinson's disease.  
XX  
XX Homo sapiens.  
XX  
XX US2002107180-A1.  
XX  
XX 08-AUG-2002.  
XX  
XX 30-NOV-2001, 2001US-0997701.  
XX  
XX PF

XX  
PR 06-NOV-1998; 98US-0187331.  
PR 22-DEC-1999; 99US-0470946.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Yue H, Corley NC, Guegler KJ, Gorgone GA, Baughn MR;  
XX WPI; 2002-690768/74.  
DR N-PSDB; ABX14169.  
XX  
XX  
PT New human cell surface glycoproteins useful for the diagnosis,  
PT treatment and prevention of hematologic, karyotypic and neuronal  
PT disorders -  
XX  
XX Claim 1, Fig 1; 42pp; English.  
PS  
XX The invention discloses the isolated human cell surface glycoproteins  
XX (CSGP), CSGP-1 or CSGP-2. Cell surface proteins play critical roles in  
XX cell-cell recognition and adhesion, cell motility and signal transduction  
XX and act as stage-specific markers of cell differentiation and as  
XX immunological identification of distinct cell types. The polypeptides can  
XX be used for screening a compound for effectiveness as an agonist or  
XX antagonist and for identifying a compound that specifically binds to or  
XX modulates their activity. The nucleic acids are useful for screening a  
XX compound for effectiveness in altering expression of a target  
XX polynucleotide or for assessing toxicity of a test compound. Antibodies  
XX against CSGP are useful in a diagnostic test for a condition or a disease  
XX associated with the expression of CSGP in a biological sample and for  
XX detecting or purifying CSGP from a sample. The CSGP compositions are  
XX useful for treating a disease or condition associated with decreased or  
XX increased expression of functional CSGP. The polypeptides, nucleic acids  
XX (gene therapy) and vaccines are useful for diagnosing, treating and  
XX preventing haematologic (e.g. anaemia, haemorrhage, thrombosis,  
XX haematopoietic disorders, haemoglobin disorders, including sickle cell  
XX anaemia, bone marrow disorders, leukaemia and other myeloproliferative  
XX disorders, lymphoma, Hodgkin's disease and haemophilia), karyotypic (e.g.  
XX Klinefelter syndrome and Turner syndrome) and neuronal disorders (e.g.  
XX Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, dementia,  
XX depression, Down's syndrome, epilepsy, Huntington's disease, multiple  
XX sclerosis Parkinson's disease, schizophrenia). The sequence presented is  
XX the human CSGP-1.  
XX  
SQ Sequence 195 AA;  
Query Match 100.0%; Score 1110; DB 23; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.6e-88;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MESWNGLPCLAFICPLMHARGRDPDLADALDDEPTKKNSDIYPRKPPYPOBNPD 60  
DB 1 MESWNGLPCLAFICPLMHARGRDPDLADALDDEPTKKNSDIYPRKPPYPOBNPD 60  
QY 61 SGNITPRPKRPPQPOGNSGSGYFNDVDRDGRYPYPRRPPRPAAGGGGGYSSYGN 120  
DB 61 SGNITPRPKRPPQPOGNSGSGYFNDVDRDGRYPYPRRPPRPAAGGGGGYSSYGN 120  
QY 121 DNTHGGRGYRPNRSGNTYGGDHSTYGNPEGNVAKIVSIYVVVTLGAAASYFKL 180  
DB 121 DNTHGGRGYRPNRSGNTYGGDHSTYGNPEGNVAKIVSIYVVVTLGAAASYFKL 180  
QY 181 NNRNRCRTHEPENV 195  
DB 181 NNRNRCRTHEPENV 195  
RESULT 3  
AAM23909  
ID AAM23909 standard; Protein, 168 AA.  
XX  
XX AAM23909;  
XX  
XX 12-OCT-2001 (first entry)  
XX  
XX DT





XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX EPI033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GENEST) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 DR N-PSDB; AAC00627.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 4702; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNA or poly(A) RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 CC  
 SQ Sequence 91 AA;  
 Query Match 44.2%; Score 491; DB 21; Length 91;  
 Best Local Similarity 98.8%; Pred. No. 3.2e-35;  
 Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MESWGLPCLAFCLFMHARGQDPDLADLDDPEPTKKPNSDIYPKXPYYQPENPD 60  
 DB 1 MESWGLPCLAFCLFMHARGQDPDLADLDDPEPTKKPNSDIYPKXPYYQPENPD 60  
 QY 61 SGNVYPRPKPRPQPPQPGNSGNSGCG 85  
 DB 61 SGNVYPRPKPRPQPPQPGNSGNSGCG 85  
 RESULT 6  
 ABG27670  
 ID ABG27670 standard; Protein; 345 AA.  
 XX  
 AC ABG27670;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #27661.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX

PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS91857.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 58029; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: This sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 345 AA;  
 Query Match 41.9%; Score 465; DB 22; Length 345;  
 Best Local Similarity 71.1%; Pred. No. 2.4e-32;  
 Matches 86; Conservative 2; Mismatches 11; Indels 22; Gaps 2;  
 QY 5 WGLPCLAFCLFMHARGQDPDLADLDDPEPTKKPNSDIYPKXPYYQPENPD 64  
 DB 103 WRRPD-FTIRLTARSGQGRDPLADLDDPEPTKKPNSDIYPKXPYYQPENPD 161  
 QY 65 IYPRPKPRPQPPQPGNSGSGYFNDVDRDGRYPRPRPPAGGGGGGYSYSGNSDTH 124  
 DB 162 -----YFNDVDRDGRYPRPRPPAGGGGGGYSYSGNSDTH 200  
 QY 125 G 125  
 DB 201 G 201  
 RESULT 7  
 ABG27669  
 ID ABG27669 standard; Protein; 91 AA.  
 XX  
 AC ABG27669;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #27660.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX MO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Dmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS91856.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX  
XX Claim 20; SEQ ID No 58028; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 91 AA:  
SQ  
Query Match 38.5%; Score 427; DB 22; Length 91;  
Best Local Similarity 98.6%; Pred. No. 1.1e-29;  
Matches 72; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 35 EPTKKNSTIYPRKPPYPOPNPDSGNITPRPRPQPOGNSGSGYFNDVDRD 94  
DB 19 KPTKKNSTIYPRKPPYPOPNPDSGNITPRPRPQPOGNSGSGYFNDVDRD 78  
QY 95 GRYPFRPRPPA 107  
DB 79 GRYPFRPRPPA 91

RESULT 8  
ABG00689  
ID ABG00689 standard; Protein; 73 AA.  
XX  
XX ABG00689;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #680.  
DE

XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensics;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX MO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Dmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS64876.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX  
XX Claim 20; SEQ ID No 31048; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 73 AA:  
SQ  
Query Match 33.5%; Score 371.5; DB 22; Length 73;  
Best Local Similarity 82.8%; Pred. No. 5.7e-25;  
Matches 72; Conservative 0; Mismatches 0; Indels 15; Gaps 1;  
QY 109 GCGGYSYSGNSDNTGRCGRYPNSRYGNTGCGDHSYTGPNBGMVAKIVSPISVVVV 168  
DB 2 GCGGYSYSGNSDNTGRCGRYPNSRYGNTGCGDHSYTGPNBGMVAKIVSPISVVVV 46  
QY 169 TLLGMAASYFKLNNRNCFTHEPENV 195  
DB 47 TLLGMAASYFKLNNRNCFTHEPENV 73

RESULT 9  
ABG03178  
ID ABG03178 standard; Protein; 62 AA.  
XX  
XX ABG03178;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX

XX DE Novel human diagnostic protein #3169.  
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX DE food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Dirmacac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS67365.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensic, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 33537; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensic, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 62 AA;  
XX  
XX Query Match 27.7%; Score 308; DB 22; Length 62;  
XX Best Local Similarity 98.1%; Pred. No. 1.5e-19;  
XX Matches 53; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 MESWNGIPCLAFCLFMHARGQDFDLADADPEPTKKPSNDIYPRKPPYYP 54  
XX Db 1 MESWNGIPCLAFCLFMHARGQDFDLADADPEPTKKPSNDIYPRKPPYLP 54  
XX  
XX RESULT 10  
XX AAB19128 ID AAB19128 standard; Protein; 169 AA.  
XX AC AAB19128;  
XX XX 19-FEB-2001 (first entry)  
XX DT  
XX XX Polypeptide isolated from lymph node stromal cells of fan -/- mice.  
XX DE

XX XX Lymph node stromal cell; fan -/- mice; inflammatory disorder;  
XX XX immune system disorder; cancer; viral infection;  
XX XX blood vessel growth; tumour necrosis factor disorder; arthritis;  
XX XX inflammatory bowel disease; fibroblast growth factor-mediated disorder;  
XX XX cardiac failure.  
XX OS Mus sp.  
XX OS  
XX PN WO200058463-A1.  
XX PD 05-OCT-2000.  
XX PF 18-FEB-2000; 2000WO-NZ00015.  
XX PR 25-MAR-1999; 99US-0276268.  
XX PR 26-AUG-1999; 99US-0383586.  
XX XX (GENE-) GENESIS RES & DEV CORP LTD.  
XX PA Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;  
XX PI Murison JG;  
XX DR WPI; 2000-664924/64.  
XX DR N-PSDB; AAB96738.  
XX XX  
XX PS Claim 1; Page 71; 75pp; English.  
XX XX  
XX CC The present sequence represents a polypeptide sequence which is  
XX CC isolated from lymph node stromal cells of fan -/- mice. The  
XX CC polynucleotides and their polypeptides are useful for treating an  
XX CC inflammatory disorder, disorder of immune system and cancer selected  
XX CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a  
XX CC viral disorder, in particular HIV infection and for modulating the  
XX CC growth of blood vessels. The polypeptides are useful for treating a  
XX CC tumour necrosis factor (TNF) mediated disorder, such as those selected  
XX CC from arthritis, inflammatory bowel disease and cardiac failure and a  
XX CC fibroblast growth factor-mediated disorder. It is also useful in assays  
XX CC to determine biological activity, to raise antibodies, to isolate  
XX CC corresponding ligands or receptors, to quantify levels of protein or  
XX CC cognate corresponding ligand or receptors, as anti-inflammatory agents,  
XX CC and in compositions for the treatment of skin, connective tissue and  
XX CC immune system diseases. The polynucleotide is useful as marker for  
XX CC tissue, as a chromosome marker or tags in the identification of a  
XX CC genetic disorder.  
XX CC  
XX SQ Sequence 169 AA;  
XX  
XX Query Match 15.9%; Score 176.5; DB 21; Length 169;  
XX Best Local Similarity 29.6%; Pred. No. 1.1e-07;  
XX Matches 59; Conservative 17; Mismatches 56; Indels 67; Gaps 10;  
XX  
XX QY 10 LAFCLFMHARGQDFDLADADPEPTKKPSNDIYPRKPPYYPQEPNDG----- 62  
XX Db 11 LAILGAARGAASDDPDLADLDP-----NMKPTPKA---PTPKPSGGFDLEDAI 59  
XX QY 63 -----GNIYPRKPPRPOPQNSGNSGTYFNDVDDGYPFRPRPRPRAG-G 109  
XX Db 60 PGGGGGAGAEKPPN--RPQDPKP-PRPHGDSG-ISDIDLADA-----AGGG 103  
XX QY 110 GGGGYSYSGNSDTHRGGRGYPNSRYGNTYGGDHSTYGNPQGNWAKIVSPVSVVVT 169  
XX Db 104 GGAAGRGSGDDEGCHGAGAEPE-----GTPG-----LVPGVAIVAA 143  
XX QY 170 LIGAAASYFLNNRNCFR 188  
XX Db 144 VAGAVSSFVAVOGRRRLCFR 162

	RESULT 11
ID	ABB72314 ABB72314 standard; Protein; 165 AA.
XX	
AC	ABB72314;
DT	04-APR-2002 (first entry)
DE	Rat protein isolated from skin cells SEQ ID NO: 638.
XX	
KW	Human; rat; mouse; skin cell; skin wound; cancer; growth defect; developmental defect; inflammatory disease; dermatological; vulnary; immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
OS	Rattus sp.
XX	
PN	WO200190357-AI.
PD	
XX	
PE	29-NOV-2001.
PP	
PR	24-MAY-2001; 2001MO-NZ00099. 24-MAY-2000; 2000US-2066SOP. 25-JUL-2000; 2000US-22132P.
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PI	
SQ	Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD, DR WPI; 2002-122020/16. DR N-PSDB; ABL34999.
PT	New polynucleotides and polypeptides encoded by the polynucleotides isolated from skin cells, useful for treating skin wounds, cancers, growth and developmental defects, inflammatory diseases, or for modulating immune responses - PT PT
PS	Claim 4; Page 400; 466pp; English.
CC	The present invention provides the protein and coding sequences of CDNA's isolated from human, murine and rat skin cell libraries. The sequences can be used in the development of therapeutic agents useful in the treatment of skin diseases, including skin wounds, cancer, growth defects, developmental defects and inflammatory diseases. The proteins have important roles in the induction of hair growth, cell proliferation and cell-cell interaction, in maintaining tissue integrity, in wound healing and in modulating immune responses. The present sequence is a polipeptide of the invention.
Sequence	165 AA;
Query Match	15.5%; Score 172.5; DB 23; Length 165;
Best Local Similarity	31.2%; Pred. No. 2.5e-07;
Matches	60; Conservative 18; Mismatches 55; Indels 59; Gaps 9,
OY	LAPICFLLHARGQGDPDLADLDDPEPKKNSDVIYPKKPYQPENPD-----S 61
Dd	IALLGVLTALAGGGDDPLRDDALEED-----TDKKTPKPETPKPKSSGFPLEEAULTG 64
OY	GGINIVP-----RPXRPPOPGRNGSGGYFNDVDNRDGRVPFRPRPPAGGGGGYS 116
Dd	GADEDPRRPGSRPKPDPKP-PQPRPDSSG-ISDRULED-----VAHGGRG--- 108
OY	TGNSDNTHTRGRCYRNRSRYGMTYYGGDHSTTYGNPEGNNVAKIVSETVVVVTTLIGAAAS 176
Dd	-----GGAGDR-----GTDAESSEGPOG-----LIIPGVAVTALAAGA VSS 146
OY	YFKLNRRNCFR 188
Dd	FVALYOKRLCFLR 158

Query Match	14.8% Score 164.5; DB 21; Length 195;
Best Local Similarity 28.3%; Pred. No. 1.5e-06;	
Matches 54; Conservative 14; Mismatches 56; Indels 67; Gaps 7;	
15 PLMEARGQDFDLADLDPEPTKXNSDIYPKRXPYY-----PPPE 57	
26 FLHGERNNNGFSDLSDALPDNE-NKKPTA-----IPKKSAGDPDFLGDAVVDGENDDPAP 80	
58 NPDSGNGIYRPKRPPQPPGNSGNSGTYFNDVRDGRYPPRRPRRPPAGGGGGGYSY 117	
81 NP-----PKPMPPNP-NAPSSGSGFSDALADG-----VSGGEGGSGSDG 120	

QY 118 GNSDNTGRGGRPNRGNTYGGDHSHTYGNPEGNNVAKIVSPIVSVVVTLLGAASY 177  
 Db 121 GSHRKEGEBEADAPG-----VIRGIVGAVVAVAGAISSF 155  
 QY 178 FKLNRRNCFR 188  
 Db 156 IAYOKKLCFK 166

RESULT 13  
 ABU04013  
 ID ABU04013 standard; Protein; 195 AA.  
 AC ABU04013;  
 XX  
 DT 29-JAN-2003 (first entry)  
 XX  
 DE Human expressed protein tag (EPT) #679.  
 XX  
 KW Translational profiling; expressed protein tag; EPT; kinase;  
 KW phosphatase; protease; protease inhibitor; transporter;  
 KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer;  
 KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;  
 KW leukemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200278524-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 DE 28-MAR-2002; 2002WO-US09671.  
 XX  
 PF 28-MAR-2001; 2001US-279495P.  
 PR 21-MAY-2001; 2001US-292544P.  
 PR 08-AUG-2001; 2001US-310801P.  
 PR 01-OCT-2001; 2001US-326370P.  
 PR 04-DEC-2001; 2001US-336780P.  
 PR 20-FEB-2002; 2002US-358985P.  
 XX  
 PA (ZYCO-) ZYCO INC.  
 XX  
 PI Chicx RM, Tomlinson AJ, Urban RG;  
 XX  
 DR WPI; 2003-040607/03.  
 XX  
 PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma  
 PT or leukemia -  
 XX  
 PS Example 2; SEQ ID No 679; 134pp; English.

The invention describes a purified polypeptide, which comprises a  
 fragment of a kinase, phosphatase, protease, protease inhibitor,  
 transporter, cytoskeletal protein, receptor or transcription factor.  
 The polypeptide is useful as an immunogenic composition for eliciting  
 in a mammal an immunogenic response directed against any of the purified  
 polypeptide. The purified polypeptide, or the antibody that binds to  
 this polypeptide, is useful for treating cancer. The polypeptide is  
 also useful for identifying compounds that binds to a naturally  
 processed class I or class II MHC-binding polypeptide. The polypeptides  
 and polynucleotides are particularly useful for treating or preventing  
 myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 lymphoma or leukemia. These are also useful for screening agents for  
 treating the above mentioned disease. This sequence represents an  
 expressed protein tag (EPT) isolated from human tissue for translational  
 profiling.  
 Note: This sequence does not appear in the printed specification but was  
 obtained in electronic format directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences.

SEQ Sequence 195 AA;  
 Query Match 14.8%; Score 164.5; DB 24; Length 195;  
 Best Local Similarity 28.3%; Pred. No. 1.5e-06;  
 Matches 54; Conservative 14; Mismatches 56; Indels 67; Gaps 7;

QY 15 FLMHARGQRPDLADLDDPEPTKKNSDIYPRKPPY-----PQPE 57  
 Db 26 FLHGERNNGGFDLSDALPDNE-NKKPTA-----IPKRPASAGDDFDLGDVVDGENDPFRPP 80  
 QY 58 NPDSGNIVPRPRPQPOPGNSGNSGVFNDVDRDGRVPRPRPPAGGGGGYSSY 117  
 Db 81 NP-----PKPMPPNP-NHPSSSGSFSDADLDG-----VSGSGKGSDDG 120  
 QY 118 GNSDNTGRGGRPNRGNTYGGDHSHTYGNPEGNNVAKIVSPIVSVVVTLLGAASY 177  
 Db 121 GSHRKEGEBEADAPG-----VIRGIVGAVVAVAGAISSF 155  
 QY 178 FKLNRRNCFR 188  
 Db 156 IAYOKKLCFK 166

RESULT 14  
 ABG11668  
 ID ABG11668 standard; Protein; 265 AA.  
 AC ABG11668;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #11659.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 DE 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Dimañac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 XX  
 DR N-PSDB; AAS75855.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 42027; 103pp; English.

This invention relates to isolated polynucleotide (I) and  
 polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome  
 mapping, and in recombinant production of (II). The  
 polynucleotides are also used in diagnostics as expressed sequence tags  
 for identifying expressed genes. (I) is useful in gene therapy techniques  
 to restore normal activity of (II) or to treat disease states involving  
 (II). (II) is useful for generating antibodies against it, detecting or  
 quantitating a polypeptide in tissue, as molecular weight markers and as  
 a food supplement. (II) and its binding partners are useful in medical  
 imaging of sites expressing (II). (I) and (II) are useful for treating  
 disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 265 AA;  
Query Match 14.6%; Score 162.5; DB 22; Length 265;  
Best Local Similarity 25.5%; Pred. No. 3.1e-06;  
Matches 59; Conservative 15; Mismatches 60; Indels 97; Gaps 9;  
OY 5 WGLPLADLCFLMH-----ARG-----ORDPLADALDP 34  
DB 35 WSLCPAPAFALHSGTVPARSGRTMAGALALLFGLGVVAAPDGFDSLALPDN 94  
OY 35 EPTKKPNSDIYKPKPPYY-----POPENPDSGANITYPRPKPPPOPO 77  
DB 95 E-NKKPTA-----IPKKPSAGDDFDLGDADVDDGNDPPRPNP-----PKPWPENPNP 140  
OY 78 GNSGSGGYFENDVDGDRYPRPRPRPAGGGGGYSGNSDNTHGGRGGRPSRYGN 137  
DB 141 -NHPSSSGSFSDADLADG-----VSGGEGKSGSDGGSHRKEGEADAPG----- 184  
OY 138 TYGGDHSTYGNBEGNMAKIVSPYVVVTLGAASYFKLNRRRCFR 188  
DB 185 -----VIRGIVGAVVAVAGAISSPIAYOKKLCFK 215  
RESULT 15  
ID ABU04008 standard; Protein; 185 AA.  
XX  
AC ABU04008;  
XX  
DT 29-JAN-2003 (first entry)  
XX  
DE Human expressed protein tag (EPT) #674.  
XX  
KW Translational profiling; expressed protein tag; EPT; kinase;  
KW phosphatase; proteinase; protease inhibitor; transporter;  
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer;  
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200278524-A2.  
XX  
PD 10-OCT-2002.  
XX  
PP 28-MAR-2002; 2002WO-US09671.  
XX  
PR 28-MAR-2001; 2001US-279495P.  
PR 21-MAY-2001; 2001US-292544P.  
PR 08-AUG-2001; 2001US-310801P.  
PR 01-OCT-2001; 2001US-326370P.  
PR 04-DEC-2001; 2001US-336780P.  
PR 20-FEB-2002; 2002US-358985P.  
XX  
PA (ZYCO-) ZYCOs INC.  
XX  
PI Chicx RM, Tomlinson AJ, Urban RG;  
XX  
DR WPI; 2003-040607/03.  
XX  
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for

PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma  
PT or leukemia -  
XX  
XX Example 2; SEQ ID No 674; 134pp; English.  
XX  
PS The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor.  
CC The polypeptide is useful as an immunogenic composition for eliciting  
CC in a mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to  
CC this polypeptide, is useful for treating cancer. The polypeptide is  
CC also useful for identifying compounds that binds to a naturally  
CC processed class I or class II MHC-binding polypeptide. The polypeptides  
CC and polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukaemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling.  
CC Note: This sequence does not appear in the printed specification but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 185 AA;  
Query Match 14.5%; Score 160.5; DB 24; Length 185;  
Best Local Similarity 28.0%; Pred. No. 3.1e-06;  
Matches 56; Conservative 16; Mismatches 57; Indels 71; Gaps 8;  
OY 10 LAFICF-----LMARQRDFDLADLDDPEPTKKPNSDIYKPKPPYY----- 53  
DB 7 LALILFGLGVTVVAAPDGFDSLALPDNE-NKKPTA-----IPKKPSAGDDFDLGDADVVD 61  
OY 54 -----POPENPDSGANITYPRPKPPPOPOGSGSGGYFENDVDGDRYPRPRPRPAG 108  
DB 62 GENDPPRPNP-----PKPWPENPNP-NHPSSSGSFSDADLADG-----VSG 101  
OY 109 GGGGYSYSGNSDNTHGGRGGRPSRYGNTYGGDHSTYGNBEGNMAKIVSPYVVVV 168  
DB 102 GEGKSGSDGGSHRKEGEADAPG-----VIRGIVGAVVV 136  
OY 169 TLGAASYFKLNRRRCFR 188  
DB 137 AVAGAISSPIAYOKKLCFK 156  
Search completed: February 18, 2004, 19:05:13  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:06:36 ; Search time 34 Seconds

(without alignments)  
1200.869 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110  
Sequence: 1 MESWGLPCLAFICFLMAR.....SYKLNRRNCRTHPEENV 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: 1: Published Applications AA:\*

2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1110	100.0	195	10	US-09-997-701-1
2	1004.5	90.5	180	10	US-09-997-701-5
3	176.5	15.9	169	9	US-09-823-038A-53
4	172.5	15.5	165	11	US-09-866-050A-638
5	164.5	14.8	155	9	US-09-925-302-717
6	160.5	14.5	165	12	US-10-331-496A-24
7	149	13.4	165	11	US-09-746-783-65
8	139	12.5	527	10	US-09-712-363-156
9	139	12.5	527	16	US-10-080-170-348
10	138	12.4	229	11	US-09-746-783-38
11	127.5	11.5	707	11	US-09-919-033-278
12	127	11.4	1046	15	US-10-156-761-10088
13	125	11.3	203	15	US-10-226-489-6
14	125	11.3	204	12	US-10-280-114-9
15	121.5	10.9	488	16	US-10-080-170-4

16	121.5	10.9	633	14	US-10-086-464-11	Sequence 11, Appl
17	121	10.9	455	12	US-10-032-585-7399	Sequence 7399, Ap
18	119	10.7	1004	10	US-09-738-626-5676	Sequence 5676, Ap
19	118.5	10.7	264	12	US-10-304-630-24	Sequence 24, Appl
20	117	10.5	922	10	US-09-855-754-6	Sequence 6, Appl
21	117	10.5	922	15	US-10-302-896-6	Sequence 6, Appl
22	117	10.5	922	15	US-10-227-353-6	Sequence 6, Appl
23	117	10.5	953	9	US-09-888-615-66	Sequence 66, Appl
24	116.5	10.5	940	12	US-10-291-172-363	Sequence 363, App
25	116.5	10.5	968	12	US-10-291-172-739	Sequence 739, App
26	115	10.4	264	12	US-10-301-488A-30	Sequence 30, Appl
27	115	10.4	264	12	US-10-301-488A-33	Sequence 33, Appl
28	115	10.4	264	12	US-10-410-907A-13	Sequence 13, Appl
29	115	10.4	264	12	US-10-346-190-80	Sequence 80, Appl
30	115	10.4	264	12	US-10-417-964A-19	Sequence 19, Appl
31	115	10.4	264	15	US-10-417-964A-32	Sequence 32, Appl
32	115	10.4	264	15	US-10-209-194-2	Sequence 2, Appl
33	114.5	10.3	643	10	US-09-801-368-236	Sequence 236, App
34	113	10.2	251	12	US-10-029-386-33860	Sequence 33860, A
35	111	10.0	312	15	US-10-156-761-10983	Sequence 10983, A
36	110.5	10.0	256	12	US-10-304-630-31	Sequence 31, Appl
37	110.5	10.0	258	12	US-10-301-488A-27	Sequence 27, Appl
38	110.5	10.0	264	12	US-10-304-630-27	Sequence 27, Appl
39	110	9.9	54	10	US-09-855-754-22	Sequence 22, Appl
40	110	9.9	54	12	US-10-302-896-22	Sequence 22, Appl
41	109	9.8	1228	11	US-09-917-384-1	Sequence 1, Appl
42	109	9.8	1228	11	US-09-917-383-1	Sequence 1, Appl
43	108.5	9.8	256	12	US-10-304-630-25	Sequence 25, Appl
44	108.5	9.8	256	12	US-10-304-630-26	Sequence 26, Appl
45	108.5	9.8	256	12	US-10-304-630-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-09-997-701-1  
Sequence 1, Application US/09997701  
Patent No. US20020107180A1  
GENERAL INFORMATION:  
APPLICANT: Yue, Henry  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina A.  
APPLICANT: Baugun, Mariah R.  
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
FILE REFERENCE: PF-0631 US  
CURRENT APPLICATION NUMBER: US/09/997,701  
CURRENT FILING DATE: 2001-11-30  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/470,946  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PERL Program  
SEQ ID NO 1  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: 2297891  
US-09-997-701-1

Query Match 100.0% Score 1110 DB 10 Length 195  
Best Local Similarity 100.0% Pred. No. 2.1e-87  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 MESWGLPCLAFICFLMARAGGQDFDLADLDPEPTKKKNSDIYKPKPPYPPQEPND 60  
DB 1 MESWGLPCLAFICFLMARAGGQDFDLADLDPEPTKKKNSDIYKPKPPYPPQEPND 60  
QY 61 SGNITYPKPKPPQPPQPSNGSGGYFNDVDRDGRYPKPPRPPAGGGGGYSYGS 120  
DB 61 SGNITYPKPKPPQPPQPSNGSGGYFNDVDRDGRYPKPPRPPAGGGGGYSYGS 120

Qy 121 DNTGRCGYRNSRYGNTYGGDHHSTYGNPEGNNVAKIVSIVSVVVTLLGAAASYFKL 180  
Db 121 DNTGRCGYRNSRYGNTYGGDHHSTYGNPEGNNVAKIVSIVSVVVTLLGAAASYFKL 180  
Qy 181 NNRNRCPTHEPENY 195  
Db 181 NNRNRCPTHEPENY 195

## RESULT 2

US-09-997-701-5  
Sequence 5, Application US/09997701  
Patent No. US20020107180A1  
GENERAL INFORMATION:  
APPLICANT: Yue, Henry  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina A.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
FILE REFERENCE: PF-0631 US  
CURRENT APPLICATION NUMBER: US/09/997,701  
CURRENT FILING DATE: 2001-11-30  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/470,946  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PERL Program  
SEQ ID NO 5  
LENGTH: 180  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE: -  
OTHER INFORMATION: g2499136  
US-09-997-701-5

Query March 90.5%; Score 1004.5; DB 10; Length 180;  
Best Local Similarity 92.3%; Pred. No. 2.1e-78;  
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

Qy 1 MESWAGLPCFLAFLCPLMHARGQDFDLADALDDEPTKPNSDIYPKRPYYPQPNPD 60  
Db 1 MESWAGLPCFLAFLCPLMHARGQDFDLADALDDEPTKPNSDIYPKRPYYPQPNPD 60  
Qy 61 SGNITPPKRPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGYSSYGS 120  
Db 61 SGNITPPKRPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGYSSYGS 120  
Qy 121 DNTGRCGYRNSRYGNTYGGDHHSTYGNPEGNNVAKIVSIVSVVVTLLGAAASYFKL 180  
Db 121 DNTGRCGYRNSRYGNTYGGDHHSTYGNPEGNNVAKIVSIVSVVVTLLGAAASYFKL 180  
Qy 181 NNRNRCPTHEPENY 195  
Db 181 NNRNRCPTHEPENY 195

## RESULT 3

US-09-823-038A-53  
Sequence 53, Application US/09823038A  
Patent No. US2002005835A1  
GENERAL INFORMATION:  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Abernethy, Nevin  
APPLICANT: Onrust, Rene  
APPLICANT: Kumble, Anand  
APPLICANT: Muirson, Greg  
TITLE OF INVENTION: Compositions Isolated From Stromal Cells  
FILE REFERENCE: 11000.1037c3  
CURRENT APPLICATION NUMBER: US/09/823,038A  
CURRENT FILING DATE: 2001-07-09  
NUMBER OF SEQ ID NOS: 61

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 53  
LENGTH: 169  
TYPE: PRT  
ORGANISM: Mouse  
US-09-823-038A-53

Query Match 15.9%; Score 176.5; DB 9; Length 169;  
Best Local Similarity 29.6%; Pred. No. 1.4e-07;  
Matches 59; Conservative 17; Mismatches 56; Indels 67; Gaps 10;

Qy 10 LAFPLCFMHARGQDFDLADALDDEPTKPNSDIYPKRPYYPQPNPD 62  
Db 11 LALLGVIALGAGDGDRLDADALDDEPTKPNSDIYPKRPYYPQPNPD 62  
Qy 63 -----GNITPPKRPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGYSS 109  
Db 60 PGGGGGAGGKPN-----RQPDPKP-PRHGDGCG-ISBSDLADA-----AGCG 103  
Qy 110 GGGGYSSYGNSDNTHGRCGYRNSRYGNTYGGDHHSTYGNPEGNNVAKIVSIVSVVVT 169  
Db 104 GAGRGSGDGGHGGAGGAEPE-----GTPQG-----LTPGVAAVVA 143  
Qy 170 LCGAASYPRLNRCFR 188  
Db 144 VAGAVSSFVAYORRLCFR 162

## RESULT 4

US-09-866-050A-638  
Sequence 638, Application US/09866050A  
Publication No. US2003004071A1  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Onrust, Rene  
APPLICANT: Muirson, James G.  
APPLICANT: Kumble, Krishanand D.  
TITLE OF INVENTION: Compositions Isolated From Skin Cells  
FILE REFERENCE: 11000.1011c4u  
CURRENT APPLICATION NUMBER: US/09/866,050A  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 725  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 638  
LENGTH: 165  
TYPE: PRT  
ORGANISM: Rat  
US-09-866-050A-638

Query Match 15.5%; Score 172.5; DB 11; Length 165;  
Best Local Similarity 31.2%; Pred. No. 2.9e-07;  
Matches 60; Conservative 18; Mismatches 55; Indels 59; Gaps 9;

Qy 10 LAFPLCFMHARGQDFDLADALDDEPTKPNSDIYPKRPYYPQPNPD-----S 61  
Db 13 LALLGVIALGAGDGDRLDADALDDEPTKPNSDIYPKRPYYPQPNPD 64  
Qy 62 GGNITP-----RKPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGYSS 116  
Db 65 GADEDPKRRGSRKPKP-PPRDSG-ISRDLAD-----VAGGGRG--- 108  
Qy 117 YGNSDNTGRCGYRNSRYGNTYGGDHHSTYGNPEGNNVAKIVSIVSVVVTLLGAAAS 176  
Db 109 -----GAGGR-----GTGASBEGQPG-----LTPGVAAVVA 146  
Qy 177 YPLNRCFR 188  
Db 147 FVAYORRLCFR 158



RESULT 5  
US-09-925-302-717  
Sequence 717, Application US/09925302  
Patent No. US20020044941A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 717  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-302-717

Query Match 14.8%; Score 164.5; DB 9; Length 195;  
Best Local Similarity 28.3%; Pred. No. 1,7e-06;  
Matches 54; Conservative 14; Mismatches 56; Indels 67; Gaps 7;  
QY 15 PLHARQRPDLADALDDEPTKPNSDIYPRKPPY-----PQPE 57  
DB 26 FLGERNNQGPDLSDALPDNE-NKKPTA-----IPKPSAGDDPDLDGAVVDGENDPRRP 80  
QY 58 NPDSGGINYPKPRPOPOGNSGSGYFNDVDRDGRPRPRPRPPGSGGGSY 117  
DB 81 NP-----PKMPNP-NHPSGSGSDADLDG-----VSGGSGGSDG 120  
QY 118 GNSDNTHGRGYPNSRYGTYGGDHSTYGNPEGNNVAKIVSPIVSVVTLGAASY 177  
DB 121 GSHRKEGEADAPG-----VIRGIVGAVVAVVAGAISSP 155  
QY 178 PKANRRNCR 188  
DB 156 IAYQKKKCPK 166

RESULT 6  
US-10-331-496A-24  
Sequence 24, Application US/10331496A  
Publication No. US20030228305A1  
GENERAL INFORMATION:  
APPLICANT: FRANTZ, GRETCHEN  
APPLICANT: HILLAN, KENNETH J.  
APPLICANT: PHILLIPS, HEIDI S.  
APPLICANT: POLAKIS, PAUL  
APPLICANT: SMITH, VICTORIA  
APPLICANT: SPENCER, SUSAN D.  
APPLICANT: WILLIAMS, P. MICKEY  
APPLICANT: WU, THOMAS D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
TITLE OF INVENTION: TREATMENT OF TUMOR  
FILE REFERENCE: P5014R1-PCT  
CURRENT APPLICATION NUMBER: US/10/331,496A  
CURRENT FILING DATE: 2002-12-30  
PRIOR APPLICATION NUMBER: US 60/345,444  
PRIOR FILING DATE: 2002-01-02  
PRIOR APPLICATION NUMBER: US 60/351,885  
PRIOR FILING DATE: 2002-01-25  
PRIOR APPLICATION NUMBER: US 60/360,066  
PRIOR FILING DATE: 2002-02-25  
PRIOR APPLICATION NUMBER: US 60/362,004  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/366,869  
PRIOR FILING DATE: 2002-03-20  
PRIOR APPLICATION NUMBER: US 60/366,284

PRIOR FILING DATE: 2002-03-21  
PRIOR APPLICATION NUMBER: US 60/368,679  
PRIOR FILING DATE: 2002-03-28  
PRIOR APPLICATION NUMBER: US 60/404,809  
PRIOR FILING DATE: 2002-08-19  
PRIOR APPLICATION NUMBER: US 60/405,645  
PRIOR FILING DATE: 2002-08-21  
NUMBER OF SEQ ID NOS: 95  
SEQ ID NO 24  
LENGTH: 185  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-331-496A-24

Query Match 14.5%; Score 160.5; DB 12; Length 185;  
Best Local Similarity 28.0%; Pred. No. 3.5e-06;  
Matches 56; Conservative 16; Mismatches 57; Indels 71; Gaps 8;  
QY 10 LAPLCF-----LHARQRPDLADALDDEPTKPNSDIYPRKPPY----- 53  
DB 7 LALLFLGLGVLAAPDGGFDLSDALPDNE-NKKPTA-----IPKPSAGDDPDLDGAVVD 61  
QY 54 -----POPNPDSGGINYPKPRPOPOGNSGSGYFNDVDRDGRPRPRPPAG 108  
DB 62 GENDPRPNP-----PKMPNP-NHPSGSGSDADLDG-----VSG 101  
QY 109 GGGGYSYGNSDNTHGRGYPNSRYGTYGGDHSTYGNPEGNNVAKIVSPIVSVV 168  
DB 102 GGGGSGSDGSHRKEGEADAPG-----VIRGIVGAVV 136  
QY 169 TLLGAASYPKANRRNCR 188  
DB 137 AVAGAISSPIAYQKKKCPK 156

RESULT 7  
US-09-746-783-65  
Sequence 65, Application US/09746783  
Publication No. US20030044935A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
McCoy, John M.  
Lavallee, Edward R.  
Racie, Lisa A.  
Treacy, Maurice  
Spaulding, Yikki  
Agostino, Michael J.  
Howes, Steven H.  
Pechtel, Kim  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
ENCODING THEM  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/746,783  
FILING DATE: 21-Dec-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Milasincic, Debra J.  
REGISTRATION NUMBER: 46,931  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400

Query Match	12.5%;	Score 139;	DB 10;	Length 527;
Best Local Similarity	29.4%;	Pred. No. 0.00075;		

!!  
!! TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
!! ENCODING THEM  
!!  
!! NUMBER OF SEQUENCES: 231  
!!  
!! CORRESPONDENCE ADDRESS:  
!! ADDRESSEE: Genetics Institute, Inc.  
!!

```

STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/746,783
FILING DATE: 21-Dec-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mlasincic, Debra J.
REGISTRATION NUMBER: 46,931
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 742-4214
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 229 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-746-783-38

Query Match      12.4%; Score 138; DB 11; Length 229;
Best Local Similarity 27.3%; Pred. No. 0.00037;
Matches 50; Conservative 24; Mismatches 63; Indels 46; Gaps 9;

QY 26 DLADLDPBPTKKNSDIYKPKRPYYQPENPDGSG---NTPPKRPQPQPGNSG 81
DB 39 DLADLADDDQDGR-----KFGIGRRERMMNVTTTKRPVTTTAPAN 80
QY 82 NSGQYFN-DV-----DRDGRYPRP RP RP RP PACGGGYSYSGNSDNTGGRGGRPNR 134
DB 81 TLGNDPLDALDDRRNDRDDGR-----RKPIAGGGG--FSDDKLEDIYGGGEYKPDKG 131
QY 135 YGN-TYGGDHSHTYGNPEGNNVAK--IVSPIVSVVVVTLGAASYPKLANRNCRTTHE 131
DB 132 KGGGRYGSN-----DDGSGVAEPGTIAGVASALAMALIGAVSYISYQKKCFESIQ 186
QY 192 PEN 194
DB 187 GLN 189

RESULT 11
US-09-919-039-278
; Sequence 278, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaeber, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 278
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1
US-09-919-039-278

```

```

Query Match      11.5%; Score 127.5; DB 11; Length 707;
Best Local Similarity 28.6%; Pred. No. 0.0099;
Matches 42; Conservative 5; Mismatches 57; Indels 43; Gaps 6;

QY 34 PEPTKPKNSDI-----YKPKRPYYQPENPDGSGN-----TTPPKRPQPQPGNSG 81
DB 147 FGTPPTPPPAVTSAPGAPPTFPSSGVPPTPQAGPPPPPAVPGPGPGPGPG 206
QY 82 NSGQYFNVDVDRDGRYP RP RP RP RP PACGGGYSYSGNSDNTGGRGGRPNRNYGNTYGG 141
DB 207 PKG-----GKPIGPKP-----GGPGUSTPGGHPKPPHKGGEPRG-----GR 245
QY 142 DHHSTY-----GNPEGNNVAKI 158
DB 246 QHHPYHQHGGPPPGPGGRREERI 272

RESULT 12
US-10-156-761-10088
; Sequence 10088, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10088
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10088

Query Match      11.4%; Score 127; DB 15; Length 1046;
Best Local Similarity 34.1%; Pred. No. 0.017;
Matches 45; Conservative 8; Mismatches 65; Indels 14; Gaps 7;

QY 28 ADALDDPBPTKKNSDIYKPKR--PPYYPQENPDGSGNIYRPKRPQ--POPNSGNSG 84
DB 244 AGAPGGFRPQAGAGQGRABRPQGGPQAGAPRTFGQMPRQAPRPG--GARPG 301
QY 85 GYFNVDVDR-DGGRYP RP RP RP RP PACGGGYSYSGNSDNTGGRGGRYP-NSRYGNTYGG 140
DB 302 G-----NRPNPQMPRPAAGRPFGGPGGGRGPGGGRPGGPGGGRPGGPGGGRPG 356
QY 141 GDHSTYGNPEG 152
DB 357 GGGGCGPAGRPGG 368

RESULT 13
US-10-226-489-6
; Sequence 6, Application US/10226489
; Publication No. US20030078390A1
; GENERAL INFORMATION:
; APPLICANT: Guido Christiaan Paesen, Patricia Anne Nuttall
; TITLE OF INVENTION: Tissue Cement
; FILE REFERENCE: 2488-1-001
; CURRENT APPLICATION NUMBER: US/10/226,489
; PRIOR FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: US/09/554,547
US-10-226-489-6

```

;; PRIOR PILING DATE: 2000-07-07  
;; NUMBER OP SEQ ID NOS: 17  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 6  
;; LENGTH: 203  
;; TYPE: PRT  
;; ORGANISM: R. appendiculatus  
US-10-226-489-6

Query Match 11.3%; Score 125; DB 15; Length 203;  
Best Local Similarity 29.1%; Pred. No. 0.0043;  
Matches 50; Conservative 10; Mismatches 64; Indels 48; Gaps 7;

QY 7 LPCAFLCFLMHARGQDPDLADALDDEPTKPNNDIYKPP-----KP 50  
DB 11 LAAPAFLLSAIHNNQSCVDAA-----PTRRP-----MSPGCGAGPCFTGIATLLRP 59  
QY 51 PYPQPPENDSGNITYPKPPRPOQPGNSGSGYFNDVDRDGRYP-----PRP 101  
DB 60 GCGGQPGGQPGQGRP-PMPPPGVPFGTSGSPQGRPGABR-PGVPFGTSGSPQGRPNA 117  
QY 102 RPRPPAGGCGGSGSYGNSDNTHGRGYPNSRYGNTYGGDHSTYGNPEGN 153  
DB 118 RPRPGVPGTPTVSSPGSS-----PGSSPGISIGTPTLGTPTGPGSS 159

RESULT 14  
US-10-280-114-9  
;; Sequence 9, Application US/10280114  
;; Publication No. US20030170257A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Trimmell, Adama Roseanne  
;; APPLICANT: Paesen, Guido Christiaan  
;; APPLICANT: Nuttall, Patricia Anne  
;; TITLE OF INVENTION: Vaccine Comprising A Tick Cement Protein  
;; FILE REFERENCE: 2488-1-007  
;; CURRENT APPLICATION NUMBER: US/10/280,114  
;; PRIOR PILING DATE: 2002-10-24  
;; PRIOR APPLICATION NUMBER: PCT/GB01/01834  
;; PRIOR PILING DATE: 2001-04-25  
;; PRIOR APPLICATION NUMBER: GB 0010068.5  
;; PRIOR PILING DATE: 2000-04-25  
;; PRIOR APPLICATION NUMBER: GB 0028606.2  
;; PRIOR PILING DATE: 2000-11-23  
;; NUMBER OP SEQ ID NOS: 17  
;; SOFTWARE: Seqwin9, version 1.02  
;; SEQ ID NO 9  
;; LENGTH: 204  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Clone 24  
US-10-280-114-9

Query Match 11.3%; Score 125; DB 12; Length 204;  
Best Local Similarity 29.1%; Pred. No. 0.0043;  
Matches 50; Conservative 10; Mismatches 64; Indels 48; Gaps 7;

QY 7 LPCAFLCFLMHARGQDPDLADALDDEPTKPNNDIYKPP-----KP 50  
DB 11 LAAPAFLLSAIHNNQSCVDAA-----PTRRP-----MSPGCGAGPCFTGIATLLRP 59  
QY 51 PYPQPPENDSGNITYPKPPRPOQPGNSGSGYFNDVDRDGRYP-----PRP 101  
DB 60 GCGGQPGGQPGQGRP-PMPPPGVPFGTSGSPQGRPGABR-PGVPFGTSGSPQGRPNA 117  
QY 102 RPRPPAGGCGGSGSYGNSDNTHGRGYPNSRYGNTYGGDHSTYGNPEGN 153  
DB 118 RPRPGVPGTPTVSSPGSS-----PGSSPGISIGTPTLGTPTGPGSS 159

RESULT 15  
US-10-080-170-4

;; Sequence 4, Application US/10080170  
;; Publication No. US20030129601A1  
;; GENERAL INFORMATION:  
;; APPLICANT: COLE, S.T.  
;; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
;; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
;; TREATMENT OF MYCOBACTERIOSES  
;; FILE REFERENCE: 03495.0218  
;; CURRENT APPLICATION NUMBER: US/10/080,170  
;; PRIOR PILING DATE: 2002-06-10  
;; PRIOR APPLICATION NUMBER: 60/270,123  
;; PRIOR PILING DATE: 2001-02-22  
;; NUMBER OF SEQ ID NOS: 652  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 488  
;; TYPE: PRT  
;; ORGANISM: Mycobacterium leprae  
US-10-080-170-4

Query Match 10.9%; Score 121.5; DB 16; Length 488;  
Best Local Similarity 26.4%; Pred. No. 0.022;  
Matches 51; Conservative 9; Mismatches 52; Indels 81; Gaps 12;

QY 33 DDEPTKPNNDIYKPPPTYPQP-----ENPDSG-----NIYPR 68  
DB 187 DPQGGDPFG-CYP-PPGSGYPPQAGHPPLHRPDGGYPPGQGYEDORAYHDQGGGYP 244  
QY 69 PKRPPQPGNSGNSG-----GYFNDVDR-----DDGRYPPRRP 103  
DB 245 PYPQPPATPGGYSQGHDOYRGSYPPGQGPYGGYDYGARPARPDSGYTTPSGFP 304  
QY 104 RPP-----AGGCGGYS-----SYGNSDNTHGR-----GYPNS-RYGN 137  
DB 305 APPEQRVAYPDGGGYDQYQHSGLGYRED--YGRQETQYVAENLPGYVAPSSGGYAE 362  
QY 138 TYGGDHSTYGNP 150  
DB 363 PAGRDY--DYGP 373

Search completed: February 18, 2004, 19:12:05  
Job time : 36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 18, 2004, 19:04:25 ; Search time 21 Seconds

(without alignments)  
392,887 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110  
1 MESWMLPCLAFICFLMHR.....SYFKANRNCFRTHPENV 195

Sequence: 1 MESWMLPCLAFICFLMHR.....SYFKANRNCFRTHPENV 195

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1110	100.0	195	3 US-09-187-331-1	Sequence 1, Appl1
2	1110	100.0	195	4 US-09-470-946-1	Sequence 1, Appl1
3	1004.5	90.5	180	3 US-09-187-331-5	Sequence 5, Appl1
4	1004.5	90.5	180	4 US-09-470-946-5	Sequence 5, Appl1
5	118.5	10.7	264	4 US-09-431-887-24	Sequence 24, Appl1
6	117	10.5	922	3 US-08-460-269C-6	Sequence 6, Appl1
7	115.5	10.4	129	4 US-09-328-352-6408	Sequence 6408, Ap
8	115	10.4	264	3 US-09-128-450-21	Sequence 21, Appl1
9	115	10.4	264	4 US-09-823-494-21	Sequence 21, Appl1
10	114.5	10.3	643	4 US-09-196-270-3	Sequence 3, Appl1
11	112	10.1	304	4 US-09-403-343B-22	Sequence 22, Appl1
12	111.5	10.0	721	4 US-09-252-991A-28293	Sequence 28293, A
13	110.5	10.0	256	4 US-09-431-887-31	Sequence 31, Appl1
14	110.5	10.0	264	4 US-09-431-887-27	Sequence 27, Appl1
15	109.5	9.9	1958	1 US-07-945-283-2	Sequence 2, Appl1
16	108.5	9.8	256	4 US-09-431-887-25	Sequence 25, Appl1
17	108.5	9.8	256	4 US-09-431-887-26	Sequence 26, Appl1
18	108.5	9.8	256	4 US-09-431-887-28	Sequence 28, Appl1
19	108.5	9.8	257	4 US-09-431-887-29	Sequence 29, Appl1
20	108.5	9.8	257	4 US-09-431-887-30	Sequence 30, Appl1
21	108.5	9.8	263	1 US-08-242-188-3	Sequence 3, Appl1
22	108.5	9.8	263	1 US-08-509-261A-3	Sequence 3, Appl1
23	108.5	9.8	263	1 US-08-660-626-9	Sequence 9, Appl1
24	108.5	9.8	263	1 US-08-692-892-3	Sequence 9, Appl1
25	108.5	9.8	263	2 US-08-713-939A-3	Sequence 23, Appl1
26	108.5	9.8	263	2 US-08-868-162A-23	Sequence 23, Appl1
27	108.5	9.8	263	3 US-09-031-168-9	Sequence 9, Appl1

28	108.5	9.8	263	3 US-09-036-579-3	Sequence 3, Appl1
29	108.5	9.8	263	4 US-09-550-374-3	Sequence 3, Appl1
30	108.5	9.8	263	4 US-09-943-906-3	Sequence 3, Appl1
31	108.5	9.8	264	4 US-09-627-218B-11	Sequence 11, Appl1
32	108	9.7	256	3 US-09-128-450-22	Sequence 22, Appl1
33	108	9.7	256	4 US-09-823-494-22	Sequence 22, Appl1
34	108	9.7	891	4 US-08-714-741-32	Sequence 32, Appl1
35	107.5	9.7	911	3 US-08-460-269C-4	Sequence 4, Appl1
36	106.5	9.6	907	3 US-08-989-299-12	Sequence 12, Appl1
37	106	9.5	369	4 US-09-252-991A-20460	Sequence 20460, A
38	105.5	9.5	538	4 US-09-554-572-12	Sequence 12, Appl1
39	105.5	9.5	1312	4 US-09-309-572-26	Sequence 26, Appl1
40	105.5	9.5	1737	4 US-09-309-572-13	Sequence 13, Appl1
41	105	9.5	303	3 US-08-818-112-92	Sequence 92, Appl1
42	105	9.5	303	4 US-08-818-111-93	Sequence 93, Appl1
43	105	9.5	303	4 US-09-056-556-92	Sequence 92, Appl1
44	105	9.5	303	4 US-09-072-596-93	Sequence 93, Appl1
45	105	9.5	466	3 US-08-526-136-13	Sequence 13, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-187-331-1
Sequence 1, Application US/09187331
Patent No. 6043056
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Gorgone, Gina A.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
FILE REFERENCE: PF-0631 US
CURRENT APPLICATION NUMBER: US/09/187,331
CURRENT FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 195
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2297891
US-09-187-331-1

Query Match      100.0%; Score 1110; DB 3; Length 195;
Best Local Similarity 100.0%; Pred. No. 3.8e-91;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MESWMLPCLAFICFLMHRGQRPDLADLDDPEFTKXNSDIYKPKPPYPPQENPD 60
DB      1 MESWMLPCLAFICFLMHRGQRPDLADLDDPEFTKXNSDIYKPKPPYPPQENPD 60
QY      61 SGNITPRKPRPQPPQPGNSGSGYFNDVDDGYPYPRPRPPAGGGGGSSYSGNS 120
DB      61 SGNITPRKPRPQPPQPGNSGSGYFNDVDDGYPYPRPRPPAGGGGGSSYSGNS 120
QY      121 DNTHGGRGPRPNRNYGNTTGSDHSTYGNPEGNMVAKISPIYSVVVTTLGAASYFKL 180
DB      121 DNTHGGRGPRPNRNYGNTTGSDHSTYGNPEGNMVAKISPIYSVVVTTLGAASYFKL 180
QY      181 NNRRNCFRTHPENV 195
DB      181 NNRRNCFRTHPENV 195

RESULT 2
US-09-470-946-1
Sequence 1, Application US/09470946
Patent No. 6358923
GENERAL INFORMATION:
```

APPLICANT: Yue, Henry  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina A.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
FILE REFERENCE: PF-0631 US  
CURRENT APPLICATION NUMBER: US/09/470,946  
EARLIER FILING DATE: 1999-12-22  
EARLIER APPLICATION NUMBER: US 09/187,331  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PERL Program  
SEQ ID NO 1  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE: -  
OTHER INFORMATION: 2297891  
US-09-470-946-1

Query Match 100.0%; Score 110; DB 4; Length 195;  
Best Local Similarity 100.0%; Pred. No. 3.8e-91;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKXNSDIYKPKPPYPOENPD 60  
DB 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKXNSDIYKPKPPYPOENPD 60  
QY 61 SGNITPRPKRPQPOPGNSGSGYFNDVDRDGRYPFRPRPPAGGGGGYSYGN 120  
DB 61 SGNITPRPKRPQPOPGNSGSGYFNDVDRDGRYPFRPRPPAGGGGGYSYGN 120  
QY 121 DNTGRCGYRPNRYGNTYGGDHSHTYGNPEGNNVAKIVSPISVVVVTLLGAASYFKL 180  
DB 121 DNTGRCGYRPNRYGNTYGGDHSHTYGNPEGNNVAKIVSPISVVVVTLLGAASYFKL 180  
QY 181 NNRNCFRTHEPENV 195  
DB 181 NNRNCFRTHEPENV 195

RESULT 3  
US-09-187-331-5  
Sequence 5, Application US/09187331  
Patent No. 6043056  
GENERAL INFORMATION:  
APPLICANT: Yue, Henry  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina A.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
FILE REFERENCE: PF-0631 US  
CURRENT APPLICATION NUMBER: US/09/187,331  
CURRENT FILING DATE: 1998-11-06  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PERL Program  
SEQ ID NO 5  
LENGTH: 180  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE: -  
OTHER INFORMATION: g2499136  
US-09-187-331-5

Query Match 90.5%; Score 1004.5; DB 3; Length 180;  
Best Local Similarity 92.3%; Pred. No. 7.8e-82;  
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKXNSDIYKPKPPYPOENPD 60  
DB 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKXNSDIYKPKPPYPOENPD 60

QY 61 SGNITPRPKRPQPOPGNSGSGYFNDVDRDGRYPFRPRPPAGGGGGYSYGN 120  
DB 61 SGNITPRPKRPQPOPGNSGSGYFNDVDRDGRYPFRPRPPAGGGGGYSYGN 120  
QY 121 DNTGRCGYRPNRYGNTYGGDHSHTYGNPEGNNVAKIVSPISVVVVTLLGAASYFKL 180  
DB 121 DNTGRCGYRPNRYGNTYGGDHSHTYGNPEGNNVAKIVSPISVVVVTLLGAASYFKL 180  
QY 181 NNRNCFRTHEPENV 195  
DB 181 NNRNCFRTHEPENV 195

RESULT 4  
US-09-470-946-5  
Sequence 5, Application US/09470946  
Patent No. 6358923  
GENERAL INFORMATION:  
APPLICANT: Yue, Henry  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina A.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
FILE REFERENCE: PF-0631 US  
CURRENT APPLICATION NUMBER: US/09/470,946  
EARLIER FILING DATE: 1999-12-22  
EARLIER APPLICATION NUMBER: US 09/187,331  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PERL Program  
SEQ ID NO 5  
LENGTH: 180  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE: -  
OTHER INFORMATION: g2499136  
US-09-470-946-5

Query Match 90.5%; Score 1004.5; DB 4; Length 180;  
Best Local Similarity 92.3%; Pred. No. 7.8e-82;  
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKXNSDIYKPKPPYPOENPD 60  
DB 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKXNSDIYKPKPPYPOENPD 60  
QY 61 SGNITPRPKRPQPOPGNSGSGYFNDVDRDGRYPFRPRPPAGGGGGYSYGN 120  
DB 61 SGNITPRPKRPQPOPGNSGSGYFNDVDRDGRYPFRPRPPAGGGGGYSYGN 120  
QY 121 DNTGRCGYRPNRYGNTYGGDHSHTYGNPEGNNVAKIVSPISVVVVTLLGAASYFKL 180  
DB 121 DNTGRCGYRPNRYGNTYGGDHSHTYGNPEGNNVAKIVSPISVVVVTLLGAASYFKL 180  
QY 181 NNRNCFRTHEPENV 195  
DB 181 NNRNCFRTHEPENV 195

RESULT 5  
US-09-431-887-24  
Sequence 24, Application US/09431887  
Patent No. 6534036  
GENERAL INFORMATION:  
APPLICANT: D-Gen Limited  
TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE  
DIAGNOSIS AND TREATMENT OF PRION DISEASE  
FILE REFERENCE: ICOT/P21952  
CURRENT APPLICATION NUMBER: US/09/431,887  
CURRENT FILING DATE: 1999-11-02  
PRIOR APPLICATION NUMBER: GB 9824091.4



RESULT 9  
US-09-823-494-21  
Sequence 21, Application US/09823494  
Patent No. 655610  
GENERAL INFORMATION:  
APPLICANT: Chesebro, Bruce W  
APPLICANT: Caughey, Byron W  
APPLICANT: Chabry, Joelle  
TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion  
FILE REFERENCE: 50121  
CURRENT APPLICATION NUMBER: US/09/823,494  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 09/128,450  
PRIOR FILING DATE: 1998-08-03  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 21  
LENGTH: 264  
TYPE: PRT  
ORGANISM: Bos taurus  
US-09-823-494-21

Query Match 10.4%; Score 115, DB 4, Length 264,  
Best Local Similarity 32.9%; Pred. No. 0.0078;  
Matches 28; Conservative 9; Mismatches 32; Indels 16; Gaps 3;  
QY 72 RPOPOPNSSGSGGFVNDVDRDGRYPRRP-----RPPGCGGCGGYSSYGNNDNTHGRG 127  
DB 25 KKRKPKPGGKMTGG-----SRYPGQSPGNNRYPPOGGGGMGQPHGGKQPHGGG 75  
QY 128 GYRPNRSRYGNTYGGDHHSTYGNPEG 152  
DB 76 WQGP---HGGGMGQPHGGGMGQPHG 97

RESULT 10  
US-09-196-270-3  
Sequence 3, Application US/09196270  
Patent No. 6500636  
GENERAL INFORMATION:  
APPLICANT: Hecht, Peter  
APPLICANT: Madden, Kevin  
TITLE OF INVENTION: CHIMERIC PRE-ACTIVATED TRANSCRIPTION  
FILE REFERENCE: 50078/004002  
CURRENT APPLICATION NUMBER: US/09/196,270  
CURRENT FILING DATE: 1998-11-19  
EARLIER APPLICATION NUMBER: 60/066,129  
EARLIER FILING DATE: 1997-11-19  
EARLIER APPLICATION NUMBER: 60/066,308  
EARLIER FILING DATE: 1997-11-21  
EARLIER APPLICATION NUMBER: 60/066,462  
EARLIER FILING DATE: 1997-11-24  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 643  
TYPE: PRT  
ORGANISM: Penicillium chrysogenum  
US-09-196-270-3

Query Match 10.3%; Score 114.5; DB 4, Length 643;  
Best Local Similarity 20.7%; Pred. No. 0.024;  
Matches 55; Conservative 20; Mismatches 58; Indels 133; Gaps 12;  
QY 34 PEPTKENSIDYPR-----KPPRY-----PQENPDG 62  
DB 155 PEPGKHDMMFPPQPRGSPATYPRSPINGQYSHAPPPQYQPHPPQAPNPHSY 214  
QY 63 GNIVRPKRPQPPQPGNSGNGSG-----YFNDVDR---DDGRYP----- 98

DB 215 GNLY-----VALSQGEGHPPYRRRGYDALNEFFGDKRQFPDPSYAAVQRL 265  
QY 99 -----PRPPA-GGGGGYSS----- 116  
DB 266 GLQALQPLPSGPAPRYQMPAPVAVGGGGGCGGAPPPQYHLPMSNVTRKNDLNI 325  
QY 117 -----YNSDNTHGRGGRPNRSRYGNTYGG-DHSTYGNDEGNMVAIV----- 159  
DB 326 DQFLBQMNTIYESDENVAAGVAGAHY--VHGGMNRTHSPPTSHRQATLLQPSA 383  
QY 160 -----SPISVVVVTLLGAAAY 177  
DB 384 PMAATASHPSVGTPTALTPSSAQST 409

RESULT 11  
US-09-403-343B-22  
Sequence 22, Application US/09403343B  
Patent No. 6555091  
GENERAL INFORMATION:  
APPLICANT: JOLIVET-RENAUD, COLETTE  
APPLICANT: PERRON, HERVE  
APPLICANT: MANDRAND, BERNARD  
TITLE OF INVENTION: POLYPEPTIDE CAPABLE OF REACTING WITH ANTIBODIES OF  
FILE REFERENCE: 104574  
CURRENT APPLICATION NUMBER: US/09/403,343B  
CURRENT FILING DATE: 1999-10-18  
PRIOR APPLICATION NUMBER: FR/97/05679  
PRIOR FILING DATE: 1997-04-29  
PRIOR APPLICATION NUMBER: FR/97/16870  
PRIOR FILING DATE: 1997-12-31  
PRIOR APPLICATION NUMBER: PCT/FR98/00870  
PRIOR FILING DATE: 1998-04-29  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 22  
LENGTH: 304  
TYPE: PRT  
ORGANISM: MSRV  
US-09-403-343B-22

Query Match 10.1%; Score 112; DB 4, Length 304;  
Best Local Similarity 31.0%; Pred. No. 0.017;  
Matches 35; Conservative 13; Mismatches 39; Indels 26; Gaps 7;  
QY 13 LCFILM-----HARGQRFDLADLDDEPTKPKPSIDYPRKPPYQPP-NDPSGGNIYP 67  
DB 73 LCLLSLAPHHQHQ-----DKQTQTPPPPP-----PPQPLTRPDANPSI--NSHN 121  
QY 68 RPKRPQPPQPN-----SGNSGGYFNDVDRDGRYP-----PRRPPRPAAGG 110  
DB 122 KPKRNEGTGDHQAQGRKRTKGDPPDPGSGVFLKPTLPPLPPPPPPGPG 174

RESULT 12  
US-09-252-991A-28293  
Sequence 28293, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 28293



```

; LENGTH: 721
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28293

Query Match
Best Local Similarity 10.0%; Score 111.5; DB 4; Length 721;
Matches 50; Conservative 16; Mismatches 63; Indels 49; Gaps 11;

QY 39 KPNSDIYKPKRPYPPO-----PENPDGNGIYPRKP-----RPOP-----75
Db 121 RPAADRSQLPQPPRPPYPPGPFPTGPTQRPQRLSPGGRAGTGPAGQPPRDLDP 180
QY 76 -QPNSSGSGYFENDVDRD--DGRYP--PRPPRPPAGGSGGYSYSGNSDNTGRCGYRP 131
Db 181 ROPHRVPSGSHLQDGP RPAPAGRRPVARARAPAPAG-----RCVDHADADH-RQYQRP 234
QY 132 NSRIGNTYGGDHSHTYG--NPEGNNVAKIVSVVVVTLGAASFKKNNRNCRR 188
Db 235 -----GDHRRRGRPPDPQAG--APCTARATT--GRAGSGRLTSPRCRR 277

RESULT 13
US-09-431-887-31
; Sequence 31, Application US/09431887
; Patent No. 6534036
; GENERAL INFORMATION:
; APPLICANT: D-Gen Limited
; TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PRION DISEASE
; FILE REFERENCE: ICOT/P21952
; CURRENT APPLICATION NUMBER: US/09/431,887
; CURRENT FILING DATE: 1999-11-02
; PRIOR APPLICATION NUMBER: GB 9824091.4
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 31
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-431-887-31

Query Match
Best Local Similarity 10.0%; Score 110.5; DB 4; Length 256;
Matches 44; Conservative 12; Mismatches 67; Indels 21; Gaps 6;

QY 70 KPRPQPGSGNSGCGYFNVDRDDGRYPRRP--RPPAGGGGCGYSSGNSDNTGRCG 127
Db 24 KKRPPGCGMNTGGGSRYPGQSGPQGNRYPPQGGGCGPHGGGCGPHGGG 83
QY 128 GYRPNRYGNTYGGDHSHTYG--NPEGNNV--AKIVSVSVVVVTLGAAS-----176
Db 84 WGPBHGQ-GWGGGSGSHSGWKPKPKTKNNKVVAGAAAAGAVVGGYVTLGSAMSRPLI 132
QY 177 -----YFKANNRNCFTTHERPENV 195
Db 143 HFGNDYEDRYRREMYR--YPEQV 164

RESULT 14
US-09-431-887-27
; Sequence 27, Application US/09431887
; Patent No. 6534036
; GENERAL INFORMATION:
; APPLICANT: D-Gen Limited
; TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PRION DISEASE
; FILE REFERENCE: ICOT/P21952
; CURRENT APPLICATION NUMBER: US/09/431,887
; CURRENT FILING DATE: 1999-11-02
; PRIOR APPLICATION NUMBER: GB 9824091.4
; PRIOR FILING DATE: 1999-11-04
```

```

; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Trigemellaphus strepsiceros
US-09-431-887-27

Query Match
Best Local Similarity 10.0%; Score 110.5; DB 4; Length 264;
Matches 27; Conservative 9; Mismatches 34; Indels 15; Gaps 3;

QY 72 RPOPQPGNSGSGYFNDVDRDDGRYPRRP--RPPAGGGGCGYSSGNSDNTGRCG 127
Db 25 KKRPPGCGMNTGG-----SRYPGQSGPQGNRYPPQGGGCGPHGGGCGPHGGG 75
QY 128 GYRPNRYGNTYGGDHSHTYGNEG 152
Db 76 GW--GQPHGGGCGPHGGGCGPHG 98

RESULT 15
US-07-945-283-2
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Weasley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EPO and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4128
; TELEFAX: 309-685-4011 ext. 513
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-945-283-2

Query Match
Best Local Similarity 9.9%; Score 109.5; DB 1; Length 1958;
Matches 42; Conservative 5; Mismatches 22; Indels 83; Gaps 9;

QY 48 PPPPYPPQPNPDGNGIY-----RPKPR-PQPOP-----77
Db 393 PESP--PRSTSSSSSHGQPPSTRPPRPPRPPRPPQKISRTBASENTAGTLFSH 450
QY 78 -----GNSGSGGYFNDVDRDDGRYPRRP-----P 105
Db 451 SSKLFSHPMGEGRDRGTAG--EGDRDPR--PSPRPPRPPPLPPPPPPPPPP 506
```

Qr 106 PAGG-----GGGYSYSGNSDNTHGRCYR 130  
Db 507 PAGGARRRRRRGGG-----GPPGRGRR 530  
Search completed: February 18, 2004, 19:07:37  
Job time : 22. secs

---